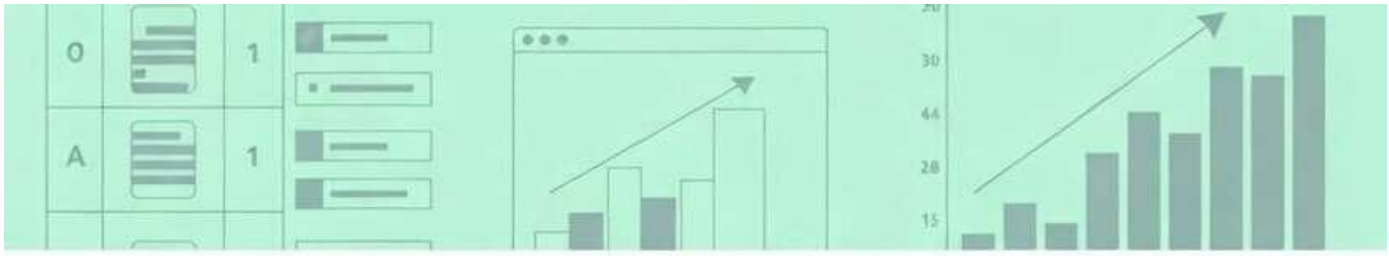


Common Statistical Test in SPSS

အသုံးများ Statistical Test များအား SPSS တွင် ဆောင်ရွက်နည်း

Dr Nay Win Aung

ဒေါက်တာနေဝင်းအောင် (ဖွံ့ဖြိုးမှုပညာ)

မာတိကာ

Frequencies (ကြိမ်နှုန်းများ / ရေတွက်မှုများ) 1
Descriptives (ဖော်ပြချက်ဆိုင်ရာ စာရင်းအင်းများ) 10
Explore (လေ့လာစူးစမ်းခြင်း) 14
Crosstabs (အပြန်အလှန်ယှဉ်တွဲဇယားများ) 20
Summarize (အကျဉ်းချုပ်ဖော်ပြခြင်း) 28
T Tests (တီ-တက်စ်များ) 33
1. Independent-Samples T Test 33
2. Paired-Samples T Test 37
3. One-Sample T Test 41
One-Way ANOVA (တစ်ဖက်သတ် ကွဲလွဲကိန်း ခွဲခြမ်းစိတ်ဖြာခြင်း) 45
Bivariate Correlations (ပြောင်းလဲကိန်းနှစ်ခုကြား ဆက်စပ်မှု) 53
Partial Correlations (တစ်စိတ်တစ်ပိုင်း ဆက်စပ်မှုများ) 59
Linear models (လီနီယာ မော်ဒယ်များ) 63
Linear Regression (မျဉ်းဖြောင့် ဆက်စပ်မှု ခွဲခြမ်းစိတ်ဖြာခြင်း) 79
Ordinal Regression (အစဉ်လိုက်ဖြစ်နေသော အမျိုးအစားခွဲ ခန့်မှန်းခြင်း) 90
Curve Estimation (မျဉ်းကွေးပုံစံ ခန့်မှန်းခြင်း) 98
Partial Least Squares Regression (တစ်စိတ်တစ်ပိုင်း အနည်းဆုံး နှစ်ထပ်ကိန်းများဖြင့် ဆက်စပ်မှု ခွဲခြမ်းစိတ်ဖြာခြင်း) 103
Ratio Statistics (အချိုးအစားဆိုင်ရာ စာရင်းအင်းများ) 108
P-P Plot (P-P ဖြစ်တန်စွမ်း ပုံပြဇယားများ) 113
Q-Q Plot (Q-Q ဖြစ်တန်စွမ်း ပုံပြဇယားများ) 118
Reliability Analysis (ယုံကြည်စိတ်ချရမှု ခွဲခြမ်းစိတ်ဖြာခြင်း) 123
ROC Curves (ROC မျဉ်းကွေးများ) 129

Frequencies (ကြိမ်နှုန်းများ / ရေတွက်မှုများ)

Frequencies procedure က ကွဲပြားတဲ့ variables အမျိုးအစားတော်တော်များများကို ဖော်ပြဖို့အတွက် အသုံးဝင်တဲ့ statistics (စာရင်းအင်းများ) နဲ့ graphical displays (ပုံပြဇယားများ) ကို ထောက်ပံ့ပေးပါတယ်။ သင့်ရဲ့ data တွေကို စတင်လေ့လာဖို့အတွက် Frequencies procedure က အကောင်းဆုံး နေရာတစ်ခုပါပဲ။ Frequency report နဲ့ bar chart တစ်ခုအတွက်၊ သင်က မတူညီတဲ့ values (distinct values) တွေကို အစဉ်လိုက် (ascending) ဒါမှမဟုတ် ပြောင်းပြန်အစဉ်လိုက် (descending) စီစဉ်နိုင်သလို၊ categories တွေကို သူတို့ရဲ့ frequencies အလိုက်လည်း စီစဉ်နိုင်ပါတယ်။ Variable တစ်ခုမှာ distinct values တွေ အရမ်းများနေတဲ့အခါ frequencies report ကို ဖျောက်ထား (suppress) လို့ ရပါတယ်။ Charts တွေမှာ frequencies ဒါမှမဟုတ် percentages (ရာခိုင်နှုန်း) တွေနဲ့ label တပ်နိုင်ပါတယ် (ပုံမှန်အားဖြင့်တော့ frequencies နဲ့ ပြပါတယ်)။

Example (ဥပမာ)။

ကုမ္ပဏီတစ်ခုရဲ့ customers တွေကို industry type (လုပ်ငန်းအမျိုးအစား) အလိုက် ဘယ်လို ဖြန့်ကျက်ထားသလဲ ဆိုတာ ကြည့်ရအောင်။ Output ကနေတစ်ဆင့် သင့် customers တွေရဲ့ 37.5% က အစိုးရအေဂျင်စီတွေ၊ 24.9% က ကော်ပိုရေးရှင်းကြီးတွေ၊ 28.1% က ပညာရေးအဖွဲ့အစည်းတွေနဲ့ 9.4% က ကျန်းမာရေးစောင့်ရှောက်မှု လုပ်ငန်းတွေက ဖြစ်တယ်ဆိုတာကို သိရှိနိုင်ပါတယ်။ Sales revenue လိုမျိုး ဆက်တိုက်ဖြစ်နေတဲ့ (continuous)၊ အရေအတွက်ပြ (quantitative) data တွေအတွက်ဆိုရင်တော့၊ average product sale (ပျမ်းမျှ အရောင်း) က $3,576 ဖြစ်ပြီး standard deviation က $1,078 ဖြစ်တယ်ဆိုတာမျိုးကို သိရှိနိုင်မှာပါ။

Statistics and plots (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များနှင့် ပုံများ)

Frequency counts (ရေတွက်မှုများ)၊ percentages (ရာခိုင်နှုန်းများ)၊ cumulative percentages (စုဆောင်းရာခိုင်နှုန်းများ)၊ mean (ပျမ်းမျှကိန်း)၊ median (အလယ်ကိန်း)၊ mode (အကြိမ်အများဆုံးကိန်း)၊ sum (ပေါင်းလဒ်)၊ standard deviation (စံသွေဖည်ကိန်း)၊ variance (ကွဲလွဲကိန်း)၊ range (အကွာအဝေး)၊ minimum နဲ့ maximum values (အနည်းဆုံးနဲ့ အများဆုံးတန်ဖိုးများ)၊ standard error of the mean (ပျမ်းမျှကိန်း၏ စံအမှား)၊ skewness နဲ့ kurtosis (သူတို့ရဲ့ standard errors တွေနဲ့အတူ)၊ quartiles၊ user ကိုယ်တိုင်သတ်မှတ်ထားတဲ့ percentiles တွေ၊ bar charts၊ pie charts၊ နဲ့ histograms တွေ ပါဝင်ပါတယ်။

Data considerations (Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ)

- **Data**

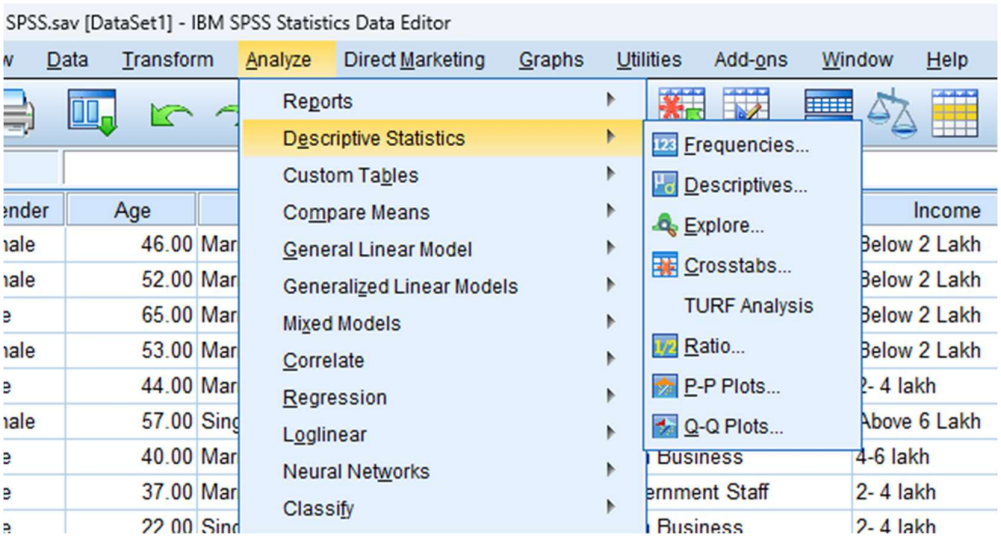
  Categorical variables (nominal သို့မဟုတ် ordinal အဆင့် တိုင်းတာမှုများ) ကို code လုပ်ဖို့အတွက် numeric codes (ဂဏန်းကုဒ်များ) ဒါမှမဟုတ် strings (စာသားများ) ကို အသုံးပြုပါ။

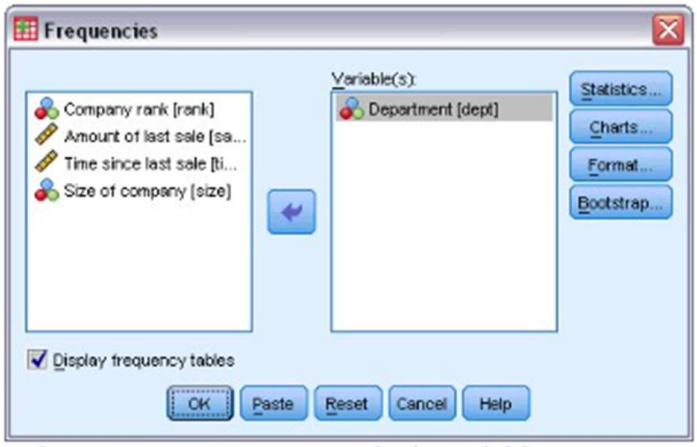
- **Assumptions (ယူဆချက်များ)**

  Tabulations နဲ့ percentages တွေက မည်သည့် distribution ကနေလာတဲ့ data အတွက်မဆို၊ အထူးသဖြင့် ordered ဒါမှမဟုတ် unordered categories တွေရှိတဲ့ variables တွေအတွက် အသုံးဝင်တဲ့ ဖော်ပြချက်ကို ပေးစွမ်းပါတယ်။ Mean နဲ့ standard deviation လိုမျိုး ရွေးချယ်နိုင်တဲ့ summary statistics အများစုဟာ normal theory အပေါ်မှာ အခြေခံထားတာဖြစ်ပြီး symmetric distributions ရှိတဲ့ quantitative variables တွေအတွက် သင့်လျော်ပါတယ်။ Median, quartiles နဲ့ percentiles လိုမျိုး robust statistics တွေကတော့ normality assumption နဲ့ ကိုက်ညီသည်ဖြစ်စေ၊ မကိုက်ညီသည်ဖြစ်စေ quantitative variables တွေအတွက် အသုံးပြုနိုင်ပါတယ်။

Obtaining frequency tables (Frequency tables များ ရယူခြင်း)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Descriptive Statistics > Frequencies...



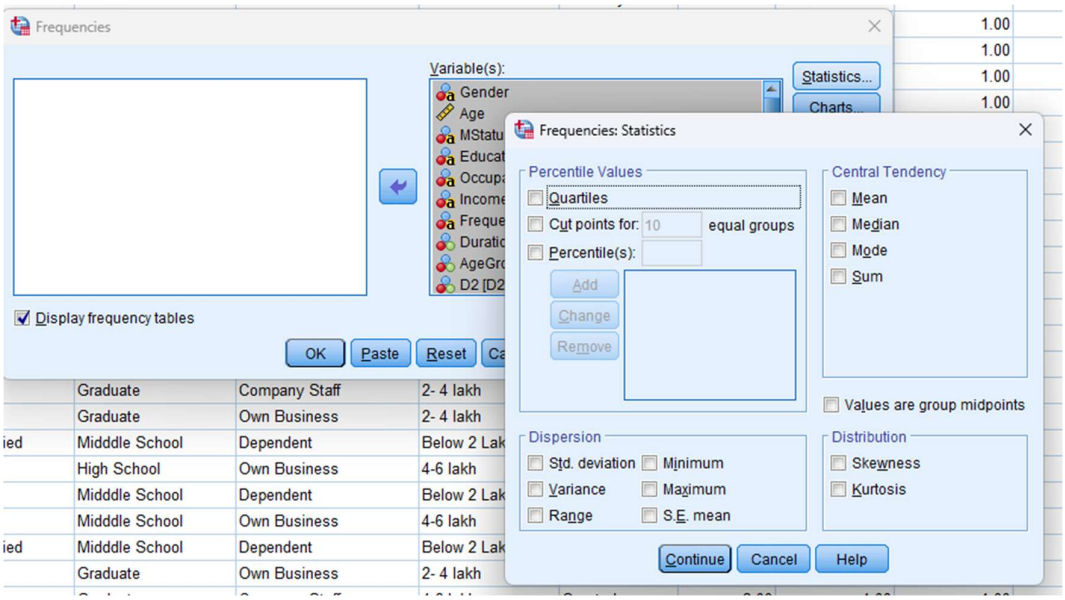


၂။ Categorical ဒါမှမဟုတ် quantitative variable တစ်ခု သို့မဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။

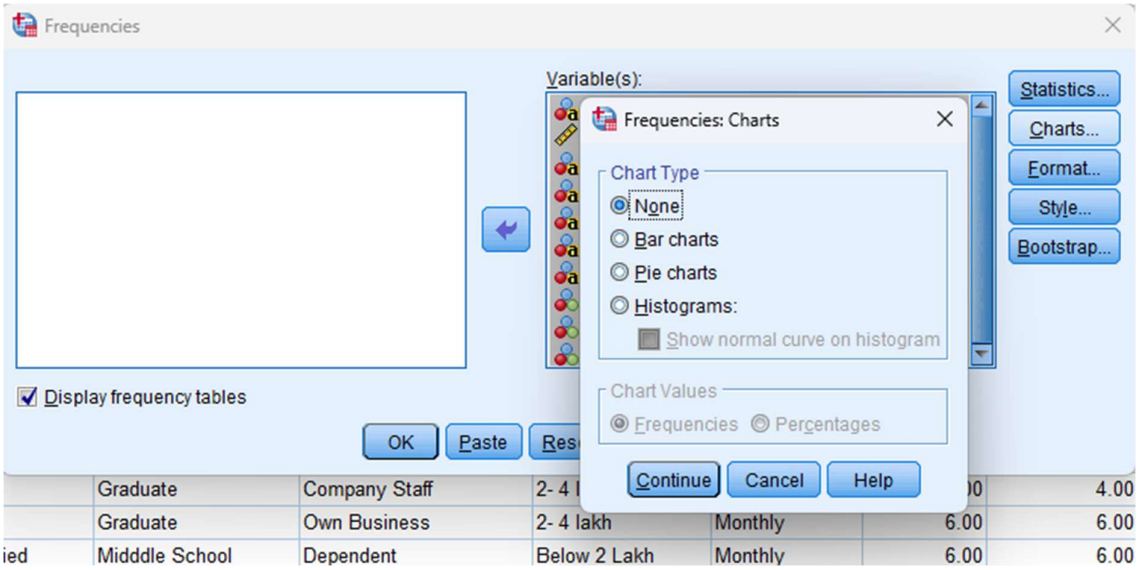
၃။ APA style လမ်းညွှန်ချက်တွေနဲ့ ကိုက်ညီတဲ့ output tables တွေ ဖန်တီးဖို့ Create APA style tables ကို ရွေးချယ်နိုင်ပါတယ် (optional)။

၄။ နောက်ထပ် ရွေးချယ်နိုင်တာတွေကတော့-

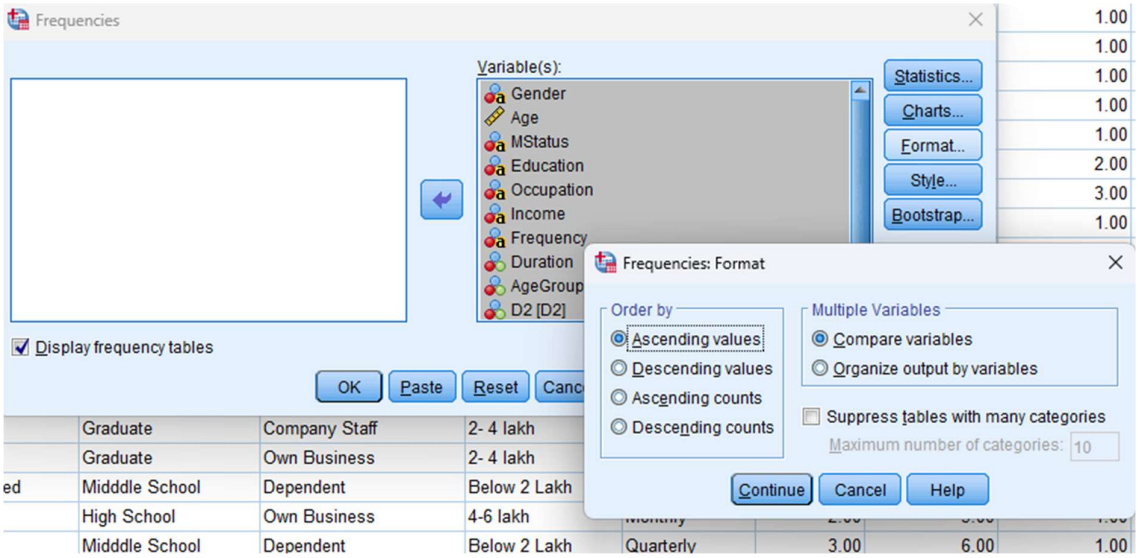
- Quantitative variables တွေအတွက် descriptive statistics ကို လိုချင်ရင် **Statistics** ကို နှိပ်ပါ။



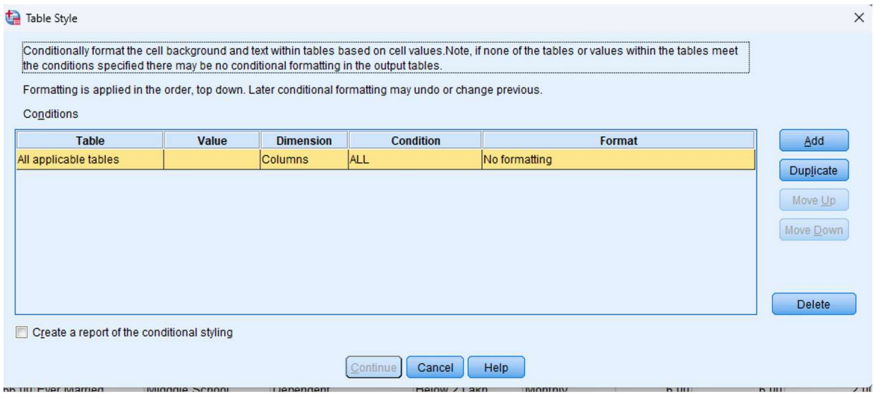
- Bar charts, pie charts နဲ့ histograms တွေအတွက် **Charts** ကို နှိပ်ပါ။



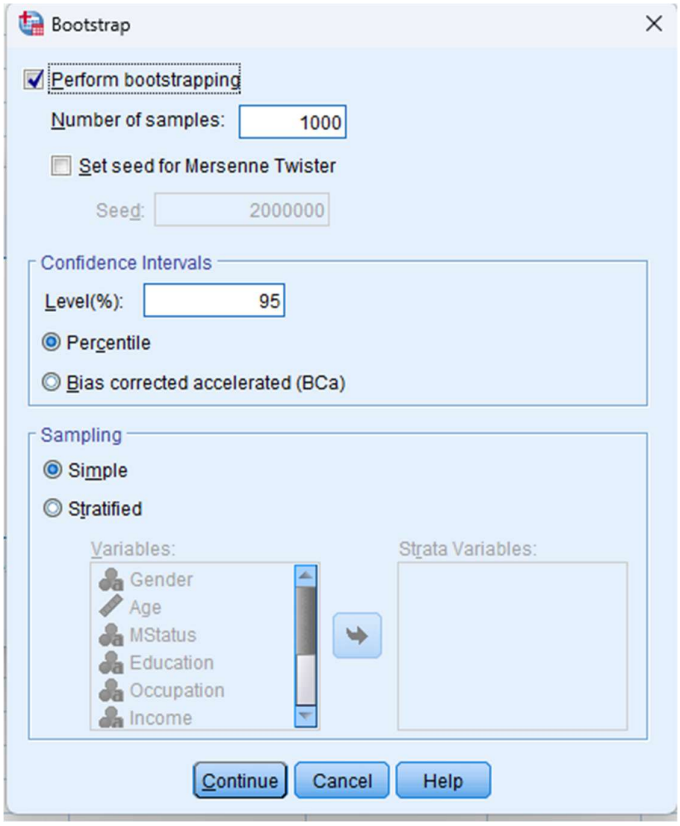
- Results တွေ ပြသမယ့် အစီအစဉ်အတွက် **Format** ကို နှိပ်ပါ။



- သီးခြား အခြေအနေတွေအပေါ် မူတည်ပြီး pivot tables တွေရဲ့ properties တွေကို အလိုအလျောက် ပြောင်းလဲပေးဖို့ **Style** ကို နှိပ်ပါ။



- Mean, median, proportion, odds ratio, correlation coefficient ဒါမှမဟုတ် regression coefficient လိုမျိုး ခန့်မှန်းချက်တွေအတွက် robust estimates of standard errors နဲ့ confidence intervals တွေကို ရယူဖို့ **Bootstrap** ကို နှိပ်ပါ။ ၎င်းကို hypothesis tests တွေ တည်ဆောက်ဖို့အတွက်လည်း အသုံးပြုနိုင်ပါတယ်။



Frequencies Statistics (Frequencies ဆိုင်ရာ စာရင်းအင်းများ)

- **Percentile Values**

  Ordered data (အစဉ်လိုက်စီထားတဲ့ data) တွေကို အထက်မှာ ရာခိုင်နှုန်းတစ်ခု၊ အောက်မှာ နောက်ရာခိုင်နှုန်းတစ်ခု ရှိနေအောင် အုပ်စုတွေခွဲပေးတဲ့ quantitative variable ရဲ့ တန်ဖိုးတွေဖြစ်ပါတယ်။ Quartiles (25th, 50th နဲ့ 75th percentiles) က observations တွေကို အရွယ်အစားတူညီတဲ့ အုပ်စုလေးခုအဖြစ် ခွဲပေးပါတယ်။ တကယ်လို့ သင်က လေးခုထက်မကတဲ့ တူညီတဲ့အုပ်စုတွေကို လိုချင်တယ်ဆိုရင် Cut points for n equal groups ကို ရွေးချယ်ပါ။ ကိုယ့်စိတ်ကြိုက် percentiles တွေကိုလည်း သတ်မှတ်နိုင်ပါတယ် (ဥပမာ- 95th percentile ဆိုရင် observations ရဲ့ 95% ဟာ အဲဒီတန်ဖိုးအောက်မှာ ရှိနေတာပါ)။

- **Central Tendency (ဗဟိုသို့တိမ်းညွတ်မှု)**

  Distribution ရဲ့ တည်နေရာကို ဖော်ပြတဲ့ statistics တွေဖြစ်ပြီး mean, median, mode နဲ့ values အားလုံးရဲ့ sum တို့ ပါဝင်ပါတယ်။

  - ***Mean***။ Central tendency ကို တိုင်းတာမှုတစ်ခုပါ။ Arithmetic average လို့လည်းခေါ်ပြီး၊ ပေါင်းလဒ်ကို cases အရေအတွက်နဲ့ စားထားတာပါ။

  - ***Median***။ Cases တွေရဲ့ တစ်ဝက်က သူ့အထက်မှာရှိပြီး ကျန်တစ်ဝက်က သူ့အောက်မှာရှိနေတဲ့ တန်ဖိုးဖြစ်ပြီး 50th percentile လို့လည်း ခေါ်ပါတယ်။ အကယ်၍ cases အရေအတွက်က စုံကိန်းဖြစ်နေမယ်ဆိုရင်၊ median ဆိုတာ data တွေကို အစဉ်လိုက်စီလိုက်တဲ့အခါ အလယ်တည့်တည့်မှာရှိတဲ့ cases နှစ်ခုရဲ့ ပျမ်းမျှတန်ဖိုး ဖြစ်ပါတယ်။ Median ဟာ Outlier (အလွန်ကြီးမားတဲ့ ဒါမှမဟုတ် အလွန်သေးငယ်တဲ့ values အနည်းငယ်) ရဲ့ သက်ရောက်မှုကို မခံရတဲ့ central tendency တိုင်းတာမှုတစ်ခု ဖြစ်ပါတယ် (Mean ကတော့ လွှမ်းမိုးမှု ခံရနိုင်ပါတယ်) ။

  - ***Mode***။ အကြိမ်အများဆုံး ဖြစ်ပေါ်နေတဲ့ value ပါ။ တကယ်လို့ values အများကြီးက အကြိမ်အများဆုံး တူညီစွာ ဖြစ်ပေါ်နေမယ်ဆိုရင် သူတို့အားလုံးက mode တွေချည်းပါပဲ။ Frequencies procedure ကတော့ ဒီလို အများကြီးဖြစ်နေတဲ့ modes တွေထဲက အငယ်ဆုံးတန်ဖိုးကိုပဲ ပြသပေးပါတယ်။

  - ***Sum***။ Missing values မဟုတ်တဲ့ cases အားလုံးရဲ့ values တွေ ပေါင်းလဒ် သို့မဟုတ် စုစုပေါင်း ဖြစ်ပါတယ်။

- **Dispersion (ပြန့်ကျဲမှု)**

  Data တွေအတွင်းမှာ ဘယ်လောက်အထိ ကွဲပြားပြန့်ကျဲနေသလဲဆိုတာကို တိုင်းတာတဲ့ statistics တွေဖြစ်ပြီး standard deviation, variance, range, minimum, maximum နဲ့ standard error of the mean တို့ ပါဝင်ပါတယ်။

  - ***Std. deviation***။ Mean ကိုဗဟိုပြုပြီး ပြန့်ကျဲနေမှုကို တိုင်းတာတာပါ။ Normal distribution တစ်ခုမှာဆိုရင် cases တွေရဲ့ 68% ဟာ mean ရဲ့ standard deviation တစ်ခုအတွင်းမှာ ရှိနေပြီး၊ 95% ဟာ standard deviations နှစ်ခုအတွင်းမှာ ရှိနေပါတယ်။

  - ***Variance***။ Mean ကိုဗဟိုပြုပြီး ပြန့်ကျဲနေမှုကို တိုင်းတာတဲ့ နောက်တစ်ခုဖြစ်ပြီး၊ mean ကနေ ကွာဟချက်တွေကို နှစ်ထပ်ကိန်းတင်ထားတဲ့ ပေါင်းလဒ်ကို cases အရေအတွက်ထက် တစ်လျော့တဲ့ကိန်းနဲ့ စားထားတာ ဖြစ်ပါတယ်။

- ***Range***။ Numeric variable တစ်ခုရဲ့ အကြီးဆုံးနဲ့ အငယ်ဆုံး values ကြားက ကွာခြားချက်ဖြစ်ပြီး maximum ထဲက minimum ကို နုတ်ထားတာပါ။
- ***Minimum***။ Numeric variable တစ်ခုရဲ့ အငယ်ဆုံး တန်ဖိုးပါ။
- ***Maximum***။ Numeric variable တစ်ခုရဲ့ အကြီးဆုံး တန်ဖိုးပါ။
- ***S. E. mean***။ တူညီတဲ့ distribution ကနေယူထားတဲ့ sample တစ်ခုနဲ့တစ်ခုကြားမှာ mean တန်ဖိုး ဘယ်လောက်အထိ ပြောင်းလဲနိုင်သလဲဆိုတာကို တိုင်းတာတာပါ။ ၎င်းကို အမှန်တကယ်ရရှိလာတဲ့ observed mean နဲ့ မှန်းဆထားတဲ့ hypothesized value တို့ကို အကြမ်းဖျင်း နှိုင်းယှဉ်ဖို့ သုံးနိုင်ပါတယ်။

- **Distribution**

Skewness နဲ့ kurtosis တွေဟာ distribution ရဲ့ ပုံသဏ္ဌာန်နဲ့ symmetry (အချိုးညီမှု) ကို ဖော်ပြတဲ့ statistics တွေဖြစ်ပါတယ်။ သူတို့ကို သူတို့ရဲ့ standard errors တွေနဲ့အတူ ပြသပေးပါတယ်။

- ***Skewness***။ Distribution တစ်ခုရဲ့ asymmetry (အချိုးမညီမှု) ကို တိုင်းတာခြင်းပါ။ Normal distribution ဟာ symmetric ဖြစ်ပြီး skewness တန်ဖိုး 0 ရှိပါတယ်။ သိသာတဲ့ positive skewness ရှိတဲ့ distribution မှာ ရှည်လျားတဲ့ ညာဘက်အမြီးပိုင်း (right tail) ရှိပြီး၊ negative skewness မှာတော့ ရှည်လျားတဲ့ ဘယ်ဘက်အမြီးပိုင်း (left tail) ရှိပါတယ်။ လမ်းညွှန်ချက်အနေနဲ့ ပြောရရင် skewness တန်ဖိုးဟာ သူ့ရဲ့ standard error ထက် နှစ်ဆကျော်နေပြီဆိုရင် symmetry ကနေ သွေဖည်နေတယ်လို့ ယူဆနိုင်ပါတယ်။
- ***Kurtosis***။ Outliers တွေ ဘယ်လောက်အထိ ရှိနေသလဲဆိုတာကို တိုင်းတာတဲ့ အတိုင်းအတာပါ။ Normal distribution တစ်ခုအတွက်ဆိုရင် kurtosis statistic ရဲ့ တန်ဖိုးက သုည ဖြစ်ပါတယ်။ Positive kurtosis က data တွေမှာ normal distribution ထက် ပိုပြီး ပြင်းထန်တဲ့ outliers တွေ ရှိနေတယ်ဆိုတာကို ပြသပြီး၊ Negative kurtosis ကတော့ normal distribution ထက် outliers ပိုနည်းတယ်ဆိုတာကို ညွှန်ပြပါတယ်။

Values are group midpoints (တန်ဖိုးများသည် အုပ်စုများ၏ အလယ်မှတ်များဖြစ်သည်)

တကယ်လို့ သင့် data ထဲက values တွေက အုပ်စုတွေရဲ့ အလယ်မှတ်တွေ ဖြစ်နေမယ်ဆိုရင် (ဥပမာ- အသက် ၃၀ ကျော်လူတွေအားလုံးကို ၃၅ လို့ code လုပ်ထားတာမျိုး)၊ မူလ ungroup လုပ်ထားတဲ့ data တွေအတွက် median နဲ့ percentiles တွေကို ခန့်မှန်းဖို့ ဒီ option ကို ရွေးချယ်ပါ။

Frequencies Charts (Frequencies ဆိုင်ရာ ပုံများ)

မှတ်ချက်- Bootstrap dialog ပေါ်မှာ Perform bootstrapping ကို ဖွင့်ထားရင် output မှာ charts တွေ ထုတ်ပေးမှာ မဟုတ်ပါဘူး။

- **Chart Type။**

Pie chart က အစိတ်အပိုင်းတွေ စုပေါင်းပြီး တစ်ခုလုံးဖြစ်လာတာကို ပြသပါတယ်။ Pie chart ရဲ့ အချပ်လေးတစ်ခုချင်းစီက grouping variable တစ်ခုနဲ့ သတ်မှတ်ထားတဲ့ အုပ်စုတစ်ခုစီကို ကိုယ်စားပြုပါတယ်။

Bar chart က distinct value ဒါမှမဟုတ် category တစ်ခုချင်းစီအတွက် ရေတွက်မှုကို သီးခြား bar လေးတွေအနေနဲ့ ပြသပေးတဲ့အတွက်၊ categories တွေကို အလွယ်တကူ မြင်သာထင်သာ နှိုင်းယှဉ်နိုင်ပါတယ်။

Histogram မှာလည်း bars တွေပါဝင်ပေမယ့် သူတို့ကို တူညီတဲ့ interval scale ပေါ်မှာ plot လုပ်ထားတာပါ။ Bar တစ်ခုစီရဲ့ အမြင့်က အဲဒီ interval အတွင်းမှာ ကျရောက်နေတဲ့ quantitative variable ရဲ့ တန်ဖိုးအရေအတွက် (count) ကို ပြသတာပါ။ Histogram က distribution ရဲ့ ပုံသဏ္ဌာန်၊ ဗဟိုနဲ့ ပြန့်ကျဲမှုကို ပြသပေးပါတယ်။ Histogram ပေါ်မှာ normal curve လေး ထပ်ဆွဲပေးထားတာက data တွေ normally distributed ဖြစ်မဖြစ် ဆုံးဖြတ်ရာမှာ ကူညီပေးပါတယ်။

- **Chart Values။**

Bar charts တွေအတွက်၊ scale axis ကို frequency counts ဒါမှမဟုတ် percentages တွေနဲ့ label တပ်နိုင်ပါတယ်။

Frequencies Format (Frequencies ဆိုင်ရာ ပုံစံချခြင်း)

- **Order by**

Frequency table ကို data တွေထဲက အမှန်တကယ် values တွေအပေါ်မူတည်ပြီး ဒါမှမဟုတ် အဲဒီ values တွေရဲ့ count (ဖြစ်ပေါ်လာတဲ့ အကြိမ်အရေအတွက်) အပေါ်မူတည်ပြီး စီစဉ်နိုင်ပါတယ်။ ပြီးတော့ ဇယားကို ascending (အစဉ်လိုက်) ဒါမှမဟုတ် descending (ပြောင်းပြန်အစဉ်လိုက်) လည်း စီစဉ်နိုင်ပါတယ်။ ဒါပေမယ့် တကယ်လို့ သင်က histogram ဒါမှမဟုတ် percentiles ကို တောင်းဆိုထားမယ်ဆိုရင်တော့၊ Frequencies က variable ကို quantitative အဖြစ် ယူဆပြီး သူ့ရဲ့ values တွေကို ascending order နဲ့ပဲ ပြသပေးမှာ ဖြစ်ပါတယ်။

- **Multiple Variables**

တကယ်လို့ သင်က variables အများကြီးအတွက် statistics tables တွေကို ထုတ်ယူမယ်ဆိုရင်၊ variables အားလုံးကို table တစ်ခုတည်းမှာ အတူတူ ပြသဖို့ (Compare variables) ဒါမှမဟုတ် variable တစ်ခုချင်းစီအတွက် statistics table သီးခြားစီ ပြသဖို့ (Organize output by variables) ရွေးချယ်နိုင်ပါတယ်။

- **Suppress tables with many categories**

ဒီ option က သတ်မှတ်ထားတဲ့ အရေအတွက်ထက် ပိုများတဲ့ values တွေပါဝင်နေတဲ့ ဇယားတွေကို မပြသဘဲ ဖျောက်ထားဖို့ ကူညီပေးပါတယ်။

Department

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Development	16	22.9	25.8	25.8
	Computer services	30	42.9	48.4	74.2
	Finance	13	18.6	21.0	95.2
	Other	3	4.3	4.8	100.0
	Total	62	88.6	100.0	
Missing	Don't know	8	11.4		
Total		70	100.0		

Descriptives (ဖော်ပြချက်ဆိုင်ရာ စာရင်းအင်းများ)

Descriptives procedure ဟာ variables (ပြောင်းလဲကိန်း) အများကြီးအတွက် univariate summary statistics (အကျဉ်းချုပ်စာရင်းအင်းများ) ကို ဇယားတစ်ခုတည်းမှာ ဖော်ပြပေးပြီး၊ standardized values (z scores) တွေကိုပါ တွက်ချက်ပေးပါတယ်။ Variables တွေကို သူတို့ရဲ့ means (ပျမ်းမျှကိန်းများ) အရွယ်အစားအလိုက် (အငယ်မှအကြီး သို့မဟုတ် အကြီးမှအငယ်) ၊ အကွာရာစဉ်အလိုက်၊ သို့မဟုတ် သင်ရွေးချယ်ခဲ့တဲ့ အစဉ်လိုက်အတိုင်း (default အနေဖြင့်) စီစဉ်နိုင်ပါတယ်။

z scores များကို အသုံးပြုခြင်း

Z scores တွေကို save လုပ်တဲ့အခါ ၎င်းတို့ကို Data Editor ထဲက data တွေမှာ variables အသစ်တွေအနေနဲ့ ပေါင်းထည့်ပေးသွားမှာဖြစ်ပြီး၊ charts တွေ၊ data listings တွေနဲ့ analyses တွေမှာ ဆက်လက် အသုံးပြုနိုင်ပါတယ်။ ဥပမာအနေနဲ့ မတူညီတဲ့ ယူနစ်တွေနဲ့ တိုင်းတာထားတဲ့ variables တွေ (ဥပမာ - လူတစ်ဦးချင်း ဝင်ငွေနဲ့ စာတတ်မြောက်မှု ရာခိုင်နှုန်း) ကို အလွယ်တကူ နှိုင်းယှဉ်နိုင်အောင် z-score transformation က ဘုံစကေး (common scale) တစ်ခုပေါ်ကို ပြောင်းလဲပေးပါတယ်။

Example (ဥပမာ)

အရောင်းဝန်ထမ်းတစ်ဦးချင်းစီရဲ့ နေ့စဉ် အရောင်းစာရင်း data တွေ (လအတော်ကြာ ကောက်ယူထားတဲ့ စာရင်း) ရှိတယ်ဆိုပါစို့။ Descriptives procedure ကိုသုံးပြီး ဝန်ထမ်းတစ်ဦးချင်းစီရဲ့ ပျမ်းမျှ နေ့စဉ်အရောင်း (average daily sales) ကို တွက်ချက်နိုင်သလို၊ ရလဒ်တွေကို အရောင်းအများဆုံးကနေ အနည်းဆုံးအထိ အစဉ်လိုက် စီစဉ်ကြည့်နိုင်ပါတယ်။

Statistics (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များ)

Sample size (နမူနာအရွယ်အစား)၊ mean (ပျမ်းမျှကိန်း)၊ minimum (အနည်းဆုံး)၊ maximum (အများဆုံး)၊ standard deviation (စံသွေဖည်ကိန်း)၊ variance (ကွဲလွဲကိန်း)၊ range (အကွာအဝေး)၊ sum (ပေါင်းလဒ်)၊ standard error of the mean (ပျမ်းမျှကိန်း၏ စံအမှား) ၊ နဲ့ kurtosis, skewness တို့ကို သူတို့ရဲ့ standard errors တွေနဲ့အတူ ရရှိနိုင်ပါတယ်။

Descriptives အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ

- **Data (အချက်အလက်)**

မှားယွင်းမှုတွေ၊ outliers တွေနဲ့ ပုံမှန်မဟုတ်တဲ့ ဖြန့်ကျက်မှုတွေ (distributional anomalies) မရှိအောင် graphical နည်းနဲ့ အရင်စစ်ဆေးပြီးသား numeric variables တွေကို အသုံးပြုသင့်ပါတယ်။ ဒီ procedure ဟာ ဖိုင်အရွယ်အစား အလွန်ကြီးမားတဲ့ (cases ထောင်ပေါင်းများစွာပါတဲ့) data တွေအတွက် အလွန်ထိရောက်မှု ရှိပါတယ်။

- **Assumptions (ယူဆချက်များ)**

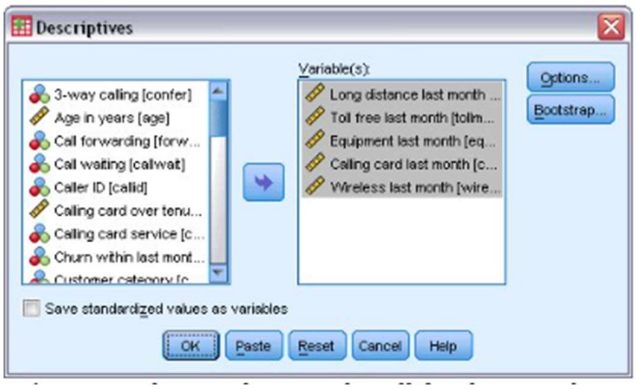
Z scores အပါအဝင် ရနိုင်တဲ့ statistics အများစုဟာ normal theory အပေါ်မှာ အခြေခံထားတာဖြစ်ပြီး symmetric distributions (အချိုးညီတဲ့ ဖြန့်ကျက်မှုများ) ရှိတဲ့ quantitative variables (အရေအတွက်ပြ ပြောင်းလဲကိန်းများ) အတွက် သင့်လျော်ပါတယ်။

အစဉ်လိုက်မရှိတဲ့ categories တွေ သို့မဟုတ် skewed distributions (တိမ်းစောင်းနေတဲ့ ဖြန့်ကျက်မှုများ) ရှိတဲ့ variables တွေကို ရှောင်ကြဉ်သင့်ပါတယ်။

Z scores ရဲ့ distribution ပုံသဏ္ဌာန်က မူရင်း data နဲ့ အတူတူပဲဖြစ်တဲ့အတွက် z scores တွက်ထုတ်လိုက်ရုံနဲ့ ပြဿနာရှိတဲ့ data ကို ဖြေရှင်းပေးနိုင်မှာ မဟုတ်ပါဘူး။

Descriptive Statistics ရယူရန် အဆင့်များ

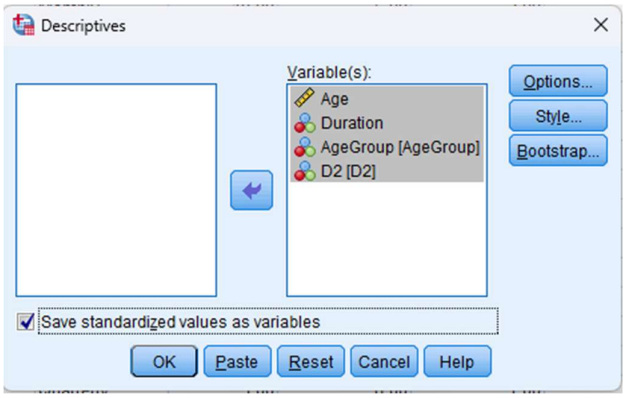
၁။ menus တွေထဲကနေ Analyze > Descriptive Statistics > Descriptives... ကို ရွေးချယ်ပါ။



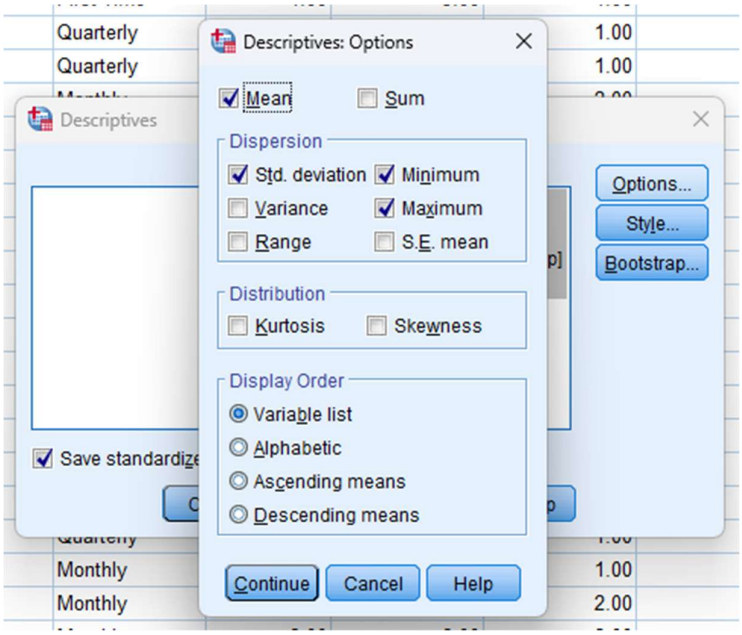
၂။ variable တစ်ခု သို့မဟုတ် အဲဒီထက်ပိုပြီး ရွေးချယ်ပါ။

ရွေးချယ်နိုင်သော ထပ်ဆောင်းအချက်များ-

- Z scores တွေကို variables အသစ်တွေအနေနဲ့ save ဖို့ **Save standardized values as variables** ကို ရွေးချယ်နိုင်ပါတယ်။



- Optional statistics တွေနဲ့ ပြသမယ့် အစဉ်လိုက် (display order) ကို ရွေးချယ်ဖို့ **Options** ကို နှိပ်နိုင်ပါတယ်။



Descriptives Options (ရွေးချယ်စရာများ)

- **Mean and Sum**

  Mean (ပျမ်းမျှကိန်း) ကို ပုံမှန်အားဖြင့် ပြသပေးထားပါတယ်။

- **Dispersion (ပြန့်ကျဲမှု)**

  Data ရဲ့ ပြန့်ကျဲမှု (spread or variation) ကို တိုင်းတာတဲ့ statistics တွေထဲမှာ standard deviation, variance, range, minimum, maximum နဲ့ standard error of the mean တို့ ပါဝင်ပါတယ်။

  - ***Std. deviation***။ Mean ကိုဗဟိုပြုပြီး ပြန့်ကျဲနေမှုကို တိုင်းတာတာပါ။ Normal distribution တစ်ခုမှာဆိုရင် cases တွေရဲ့ 68% ဟာ mean ရဲ့ standard deviation တစ်ခုအတွင်းမှာ ရှိနေပြီး၊ 95% ဟာ standard deviations နှစ်ခုအတွင်းမှာ ရှိနေပါတယ်။

  - ***Variance***။ Mean ကိုဗဟိုပြုပြီး ပြန့်ကျဲနေမှုကို တိုင်းတာတဲ့ နောက်တစ်ခုဖြစ်ပြီး၊ mean ကနေ ကွာဟချက်တွေကို နှစ်ထပ်ကိန်းတင်ထားတဲ့ ပေါင်းလဒ်ကို cases အရေအတွက်ထက် တစ်လျော့တဲ့ကိန်းနဲ့ စားထားတာ ဖြစ်ပါတယ်။

  - ***Range***။ Numeric variable တစ်ခုရဲ့ အကြီးဆုံးနဲ့ အငယ်ဆုံး values ကြားက ကွာခြားချက်ဖြစ်ပြီး maximum ထဲက minimum ကို နုတ်ထားတာပါ။

- ***Minimum / Maximum***။ အငယ်ဆုံး နဲ့ အကြီးဆုံး တန်ဖိုးတွေ ဖြစ်ပါတယ်။
- ***S.E. mean***။ Sample တစ်ခုနဲ့တစ်ခုကြားမှာ mean တန်ဖိုး ဘယ်လောက်အထိ ပြောင်းလဲနိုင်သလဲဆိုတာကို တိုင်းတာတာပါ။

- **Distribution (ဖြန့်ကျက်မှု)**

ဖြန့်ကျက်မှုရဲ့ ပုံသဏ္ဌာန်နဲ့ အချိုးညီမှုကို ဖော်ပြတဲ့ Kurtosis နဲ့ Skewness တွေကို သူတို့ရဲ့ standard errors တွေနဲ့အတူ ပြသပေးပါတယ်။

- ***Kurtosis***။ Outliers တွေ ဘယ်လောက်အထိ ရှိနေသလဲဆိုတာကို တိုင်းတာတဲ့ အတိုင်းအတာပါ။ Normal distribution တစ်ခုအတွက်ဆိုရင် သုညဖြစ်ပြီး၊ Positive က outliers ပိုများတာ၊ Negative က outliers ပိုနည်းတာကို ပြပါတယ်။
- ***Skewness***။ Distribution ရဲ့ asymmetry (အချိုးမညီမှု) ကို တိုင်းတာခြင်းပါ။ Positive skewness ရှိရင် ညာဘက်အမြီးပိုင်း ရှည်ပြီး၊ negative skewness ရှိရင် ဘယ်ဘက်အမြီးပိုင်း ရှည်ပါတယ်။

- **Display Order (ပြသမှု အစီအစဉ်)**

ပုံမှန်အားဖြင့် ရွေးချယ်ခဲ့တဲ့ အစဉ်လိုက်အတိုင်း ပြသပေမယ့်၊ အက္ခရာစဉ်အလိုက်၊ ascending means (ပျမ်းမျှကိန်း အငယ်မှအကြီး) ဒါမှမဟုတ် descending means (ပျမ်းမျှကိန်း အကြီးမှအငယ်) အတိုင်းလည်း ပြောင်းလဲပြသနိုင်ပါတယ်။

DESCRIPTIVES Command Additional Features (အခြား လုပ်ဆောင်ချက်များ)

Command syntax language သုံးပြီး အောက်ပါတို့ကိုလည်း လုပ်ဆောင်နိုင်ပါသေးတယ်-

- Variables အားလုံးမဟုတ်ဘဲ တချို့အတွက်ပဲ z scores တွေကို save လုပ်တာ (VARIABLES subcommand သုံး၍) ။
- Z scores တွေပါဝင်မယ့် variables အသစ်တွေအတွက် နာမည်သတ်မှတ်ပေးတာ။
- Missing values ပါတဲ့ cases တွေကို analysis ကနေ ဖယ်ထုတ်တာ (MISSING subcommand သုံး၍) ။
- Mean သာမက တခြား statistic တစ်ခုခုရဲ့ တန်ဖိုးအလိုက် variables တွေကို အစဉ်လိုက်စီတာ (SORT subcommand သုံး၍) ။

Explore (လေ့လာစူးစမ်းခြင်း)

Explore procedure က သင့်ရဲ့ cases တွေအားလုံးအတွက်ဖြစ်စေ၊ သီးခြား groups လေးတွေခွဲပြီးတော့ဖြစ်စေ summary statistics (အကျဉ်းချုပ် စာရင်းအင်းများ) တွေနဲ့ graphical displays (ပုံပြဇယားများ) ကို ထုတ်ပေးပါတယ်။ ဒီ Explore procedure ကို data တွေကို စစ်ဆေးတာ (data screening)၊ ထူးခြားနေတဲ့ တန်ဖိုးတွေကို ရှာဖွေတာ (outlier identification)၊ ဖော်ပြတာ (description)၊ ယူဆချက်တွေကို စစ်ဆေးတာ (assumption checking) နဲ့ subpopulations (cases အုပ်စုလေးတွေ) ကြားက ကွာခြားချက်တွေကို ဖော်ပြတာစတဲ့ အကြောင်းပြချက် အမျိုးမျိုးအတွက် အသုံးပြုကြပါတယ်။

Data တွေကို စစ်ဆေးတဲ့အခါ သင့် data ထဲမှာ ထူးခြားတဲ့တန်ဖိုးတွေ၊ လွန်ကဲနေတဲ့တန်ဖိုးတွေ (extreme values)၊ data ကွက်လပ်တွေ ဒါမှမဟုတ် တခြား ပုံမှန်မဟုတ်တဲ့အရာတွေ ရှိနေသလားဆိုတာကို သိနိုင်ပါတယ်။ Data တွေကို explore လုပ်တာက သင်သုံးမယ့် statistical techniques တွေဟာ သင့် data နဲ့ ကိုက်ညီမှု ရှိမရှိ ဆုံးဖြတ်ဖို့လည်း ကူညီပေးပါတယ်။ ဥပမာ၊ သင်သုံးမယ့် technique က normal distribution ဖြစ်ဖို့ လိုအပ်နေမယ်ဆိုရင် data ကို transform လုပ်ဖို့ လိုအပ်ကြောင်း ဒီအဆင့်မှာ တွေ့ရှိနိုင်သလို၊ nonparametric tests တွေကိုပဲ သုံးဖို့ ဆုံးဖြတ်တာမျိုးလည်း လုပ်နိုင်ပါတယ်။

Example (ဥပမာ)

မတူညီတဲ့ reinforcement schedules လေးခုအောက်မှာ ကြွက်တွေ ဝင်္ကပါ (maze) ကို လေ့လာသင်ယူတဲ့ အချိန်ကြာမြင့်မှု (maze-learning times) ကို လေ့လာကြည့်မယ်ဆိုပါစို့။ Group လေးခုထဲက တစ်ခုချင်းစီအတွက် အချိန် ဖြန့်ကျက်မှု (distribution) ဟာ normal distribution နီးပါး ဖြစ်မဖြစ်နဲ့၊ သူတို့ရဲ့ variances လေးခုစလုံး တူညီမှု ရှိမရှိကို ကြည့်ရှုနိုင်ပါတယ်။ ဒါ့အပြင် အများဆုံး အချိန် ၅ ခုနဲ့ အနည်းဆုံး အချိန် ၅ ခု ရှိတဲ့ cases တွေကိုလည်း ရှာဖွေနိုင်ပါတယ်။ Boxplots တွေနဲ့ stem-and-leaf plots တွေက groups တစ်ခုချင်းစီအတွက် သင်ယူမှုအချိန် ဖြန့်ကျက်ပုံကို ပုံနဲ့တကွ အကျဉ်းချုပ် ဖော်ပြပေးမှာ ဖြစ်ပါတယ်။

Statistics and plots (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များနှင့် ပုံများ)

Mean, median, 5% trimmed mean, standard error, variance, standard deviation, minimum, maximum, range, interquartile range, skewness နဲ့ kurtosis (သူတို့ရဲ့ standard errors တွေနဲ့အတူ)၊ mean အတွက် confidence interval (သတ်မှတ်ထားတဲ့ confidence level နဲ့)၊ percentiles၊ Huber's M-estimator, Andrews' wave estimator, Hampel's redescending M-estimator, Tukey's biweight estimator တွေ ပါဝင်ပါတယ်။ ဒါ့အပြင် အကြီးဆုံးတန်ဖိုး ၅ ခုနဲ့ အငယ်ဆုံးတန်ဖိုး ၅ ခု၊ normality ကို စစ်ဆေးဖို့အတွက် Lilliefors significance level နဲ့တကွ Kolmogorov-Smirnov statistic နဲ့ Shapiro-Wilk statistic တို့ကိုလည်း ထုတ်ပေးပါတယ်။ Plots

တွေအနေနဲ့ Boxplots, stem-and-leaf plots, histograms, normality plots နဲ့ Levene tests တွေ၊ transformations တွေပါဝင်တဲ့ spread-versus-level plots တွေကို ရရှိနိုင်ပါတယ်။

Explore Data Considerations (Explore အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစား ရမယ့် အချက်များ)

- **Data (အချက်အလက်)**

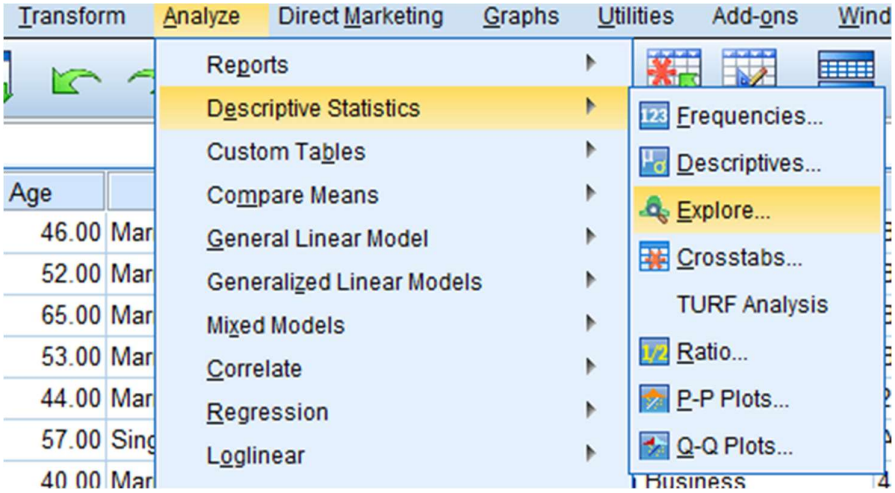
Explore procedure ကို အရေအတွက်ပြ variables (interval သို့မဟုတ် ratio-level measurements တွေဖြစ်တဲ့ quantitative variables) တွေအတွက် အသုံးပြုနိုင်ပါတယ်။ Data တွေကို အုပ်စုခွဲဖို့သုံးမယ့် factor variable မှာ သင့်လျော်တဲ့ categories အရေအတွက် ပါဝင်သင့်ပါတယ်။ ဒီတန်ဖိုးတွေက short string ဒါမှမဟုတ် numeric ဖြစ်နိုင်ပါတယ်။ Boxplots တွေမှာ outliers တွေကို label တပ်ဖို့သုံးမယ့် case label variable ကတော့ short string, long string (ပထမ ၁၅ ဘိုက်) ဒါမှမဟုတ် numeric ဖြစ်နိုင်ပါတယ်။

- **Assumptions (ယူဆချက်များ)**

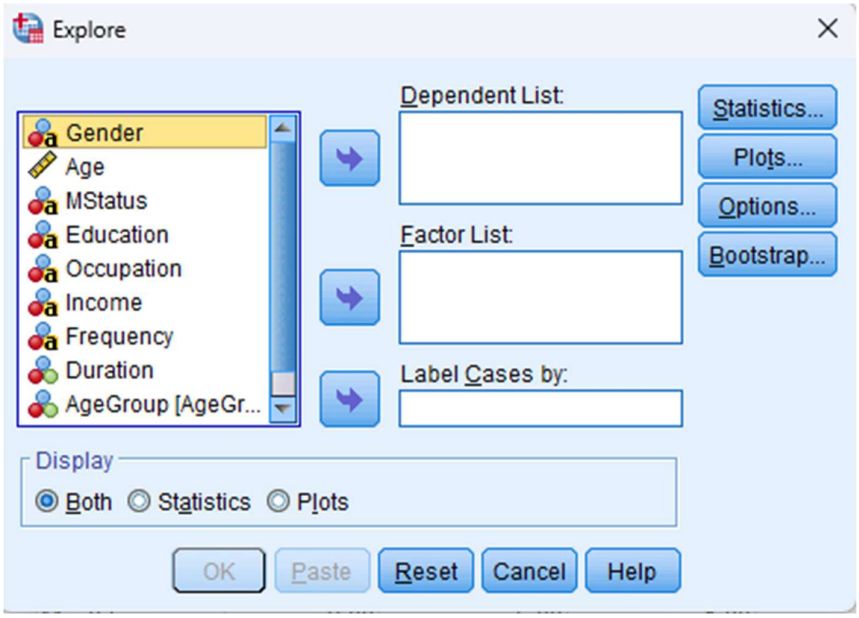
သင့် data ရဲ့ distribution ဟာ symmetric (အချိုးညီ) ဒါမှမဟုတ် normal ဖြစ်နေဖို့ မလိုအပ်ပါဘူး။

To Explore Your Data (Data များကို Explore လုပ်ရန် အဆင့်များ)

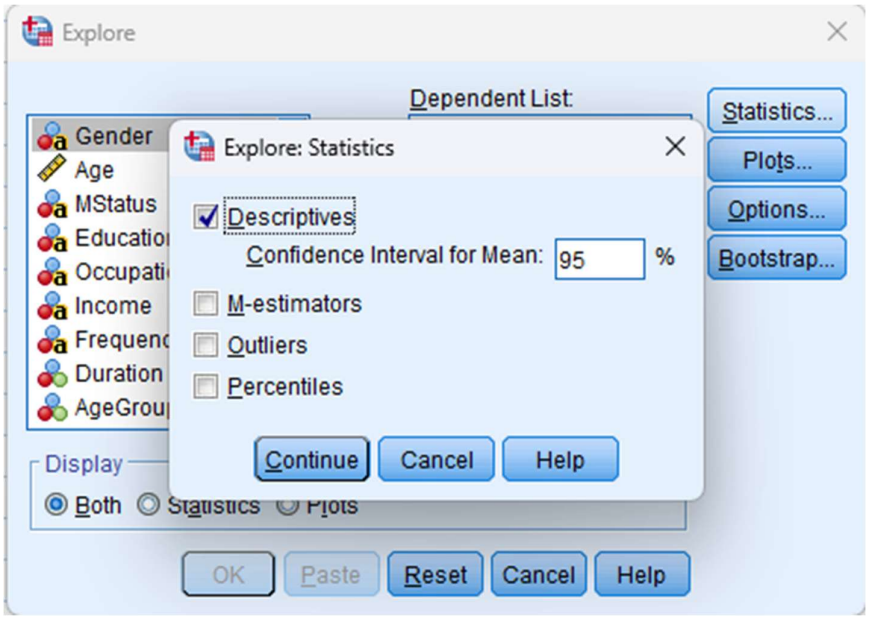
၁။ menus တွေထဲကနေ Analyze > Descriptive Statistics > Explore... ကို ရွေးချယ်ပါ။



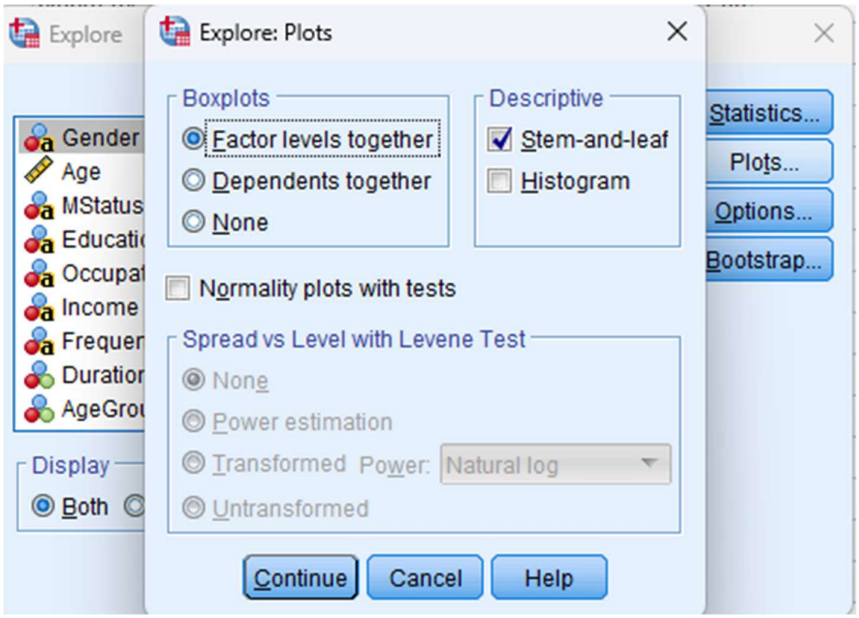
၂။ dependent variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။ ရွေးချယ်နိုင်သော ထပ်ဆောင်းအချက်များ-



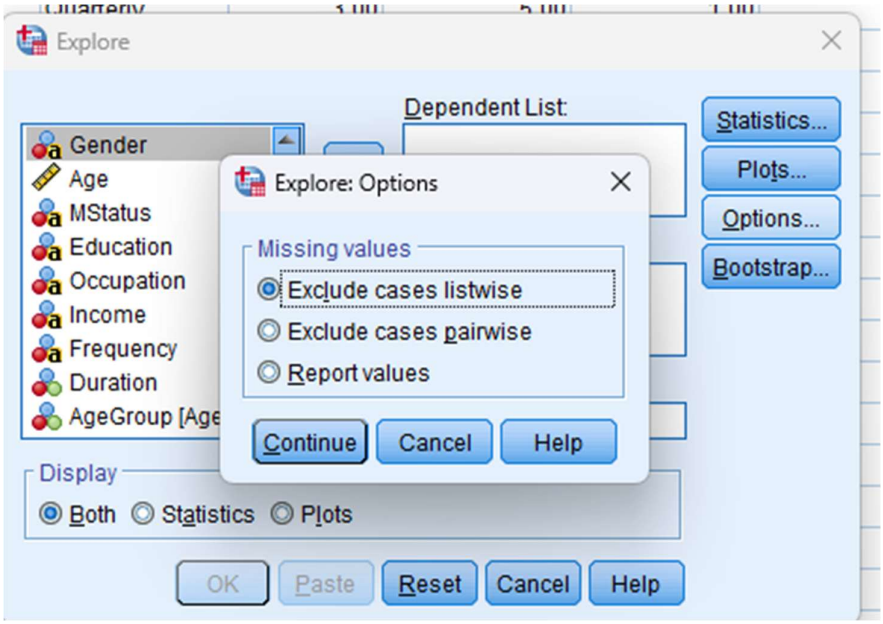
- Cases တွေကို အုပ်စုခွဲဖို့ factor variable တစ်ခု သို့မဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်နိုင်ပါတယ်။
- Cases တွေကို အမည်တပ်ဖို့ identification variable တစ်ခု ရွေးချယ်နိုင်ပါတယ်။
- Robust estimators တွေ၊ outliers တွေ၊ percentiles တွေနဲ့ frequency tables တွေအတွက် **Statistics** ကို နှိပ်ပါ။



- Histograms တွေ၊ normal probability plots နဲ့ tests တွေ၊ Levene's statistics ပါတဲ့ spread-versus-level plots တွေအတွက် **Plots** ကို နှိပ်ပါ။



- Missing values (ပျောက်ဆုံးနေတဲ့ တန်ဖိုးများ) ကို ဘယ်လိုကိုင်တွယ်မလဲဆိုတာ သတ်မှတ်ဖို့ **Options** ကို နှိပ်ပါ။



Explore Statistics (Explore ဆိုင်ရာ စာရင်းအင်းများ)

- **Descriptives**

Central tendency (ဗဟိုသို့တိမ်းညွတ်မှု) နဲ့ dispersion (ပြန့်ကျဲမှု) တိုင်းတာချက်တွေကို ပုံမှန်အားဖြင့် ပြသပေးပါတယ်။ Central tendency မှာ mean, median နဲ့ 5% trimmed mean တို့ ပါဝင်ပြီး Dispersion မှာ standard error, variance, standard deviation, minimum, maximum, range နဲ့ interquartile range တို့ ပါဝင်ပါတယ်။ Distribution ရဲ့ ပုံသဏ္ဌာန်ကို ပြတဲ့ skewness နဲ့

kurtosis တွေကိုလည်း တွေ့နိုင်ပါတယ်။ Mean အတွက် 95% confidence interval ကိုလည်း ပြသပေးပြီး အခြား confidence level ကိုလည်း ပြောင်းလဲသတ်မှတ်နိုင်ပါတယ်။

- **M-estimators**

Sample mean နဲ့ median တို့အစား အသုံးပြုနိုင်တဲ့ robust တိုင်းတာချက်တွေဖြစ်ပါတယ်။ တွက်ချက်ထားတဲ့ estimators တွေဟာ cases တွေအပေါ် သတ်မှတ်ပေးတဲ့ အလေးချိန် (weights) တွေအပေါ် မူတည်ပြီး ကွာခြားပါတယ်။

- **Outliers**

အကြီးဆုံးတန်ဖိုး ၅ ခုနဲ့ အငယ်ဆုံးတန်ဖိုး ၅ ခုကို case labels တွေနဲ့တကွ ပြသပေးပါတယ်။

- **Percentiles**

5th, 10th, 25th, 50th, 75th, 90th, နဲ့ 95th percentiles တွေအတွက် တန်ဖိုးတွေကို ပြသပေးပါတယ်။

Explore Plots (Explore ဆိုင်ရာ ပုံများ)

- **Boxplots**

Dependent variables တစ်ခုထက်ပိုရှိတဲ့အခါ boxplots တွေ ပြသပုံကို ထိန်းချုပ်နိုင်ပါတယ်။ *Factor levels together* က dependent variable တစ်ခုချင်းစီအတွက် သီးခြား display ပြသပေးပြီး၊ *Dependents together* ကတော့ factor variable က သတ်မှတ်ပေးထားတဲ့ အုပ်စုတစ်ခုချင်းစီအတွက် dependent variables တွေကို ယှဉ်လျက် ပြသပေးပါတယ်။

- **Descriptive**

Stem-and-leaf plots တွေနဲ့ histograms တွေကို ရွေးချယ်နိုင်ပါတယ်။

- **Normality plots with tests**

Normal probability နဲ့ detrended normal probability plots တွေကို ပြသပေးပါတယ်။ Normality ကို စစ်ဆေးဖို့ Lilliefors significance level နဲ့အတူ Kolmogorov-Smirnov statistic ကို ပြသပေးပါတယ်။

- **Spread vs. Level with Levene Test**

Spread-versus-level plots တွေအတွက် data transformation ကို ထိန်းချုပ်ပါတယ်။ Spread-versus-level plots အားလုံးအတွက် regression line ရဲ့ slope နဲ့ variance တူညီမှုရှိမရှိကိုစစ်ဆေးပေးတဲ့ Levene's robust tests တွေကို ပြသပေးပါတယ်။

Explore Options (Explore ဆိုင်ရာ ရွေးချယ်စရာများ) Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)။

- **Exclude cases listwise**

Dependent သို့မဟုတ် factor variable တစ်ခုခုမှာ missing value ရှိနေတဲ့ cases တွေကို analysis အားလုံးကနေ ဖယ်ထုတ်လိုက်ပါတယ် (ဒါက default ပါ)။

- **Exclude cases pairwise**

သီးခြား အုပ်စုတစ်ခု (cell) အတွက် variables တွေမှာ missing value မရှိတဲ့ cases တွေကို အဲဒီအုပ်စုရဲ့ analysis မှာ ထည့်သွင်းအသုံးပြုပါတယ်။ (အခြားအုပ်စုတွေအတွက် သုံးမယ့် variables တွေမှာတော့ missing values ရှိနေနိုင်ပါတယ်)။

- **Report values**

Factor variables တွေမှာရှိတဲ့ missing values တွေကို သီးခြား category တစ်ခုအနေနဲ့ သတ်မှတ်ပြီး ဖော်ပြပေးပါတယ်။

EXAMINE Command Additional Features (EXAMINE Command ၏ အခြားလုပ်ဆောင်ချက်များ)

Command syntax language ကိုအသုံးပြုပြီး အောက်ပါတို့ကိုလည်း လုပ်ဆောင်နိုင်ပါသေးတယ်-

- Boxplots တွေအတွက် common scale သတ်မှတ်ပေးတာ (SCALE subcommand သုံးပြီး) ။
- Default အနေနဲ့မပါတဲ့ အခြား percentiles တွေကို သတ်မှတ်တာ (PERCENTILES subcommand သုံးပြီး) ။
- M-estimators တွေအတွက် parameters တွေကို သတ်မှတ်တာ (MESTIMATORS subcommand သုံးပြီး) စတာတွေကို ထပ်မံလုပ်ဆောင်နိုင်ပါတယ်။

Crosstabs (အပြန်အလှန်ယှဉ်တွဲဇယားများ)

Crosstabs procedure ဟာ two-way (နှစ်ဖက်မြင်) နဲ့ multiway (အဖက်ဖက်မြင်) ဇယားတွေကို ဖန်တီးပေးပြီး၊ two-way tables တွေအတွက် tests တွေနဲ့ measures of association (ဆက်စပ်မှု တိုင်းတာချက်များ) မျိုးစုံကို ထောက်ပံ့ပေးပါတယ်။ ဇယားရဲ့ ဖွဲ့စည်းပုံနဲ့ categories တွေက အစဉ်လိုက်စီထားသလား (ordered) ဆိုတဲ့ အချက်တွေအပေါ် မူတည်ပြီး ဘယ် test ဒါမှမဟုတ် ဘယ် measure ကို သုံးရမလဲဆိုတာ ဆုံးဖြတ်ရတာပါ။

Partial gamma coefficients ကလွဲရင်၊ Crosstabs ရဲ့ statistics တွေနဲ့ measures of association တွေကို two-way table တစ်ခုချင်းစီအတွက် သီးခြားစီ တွက်ချက်ပေးပါတယ်။ တကယ်လို့ သင်က row တစ်ခု၊ column တစ်ခုနဲ့ layer factor (control variable) တစ်ခုကို သတ်မှတ်ပေးလိုက်မယ်ဆိုရင်၊ Crosstabs procedure ဟာ အဲဒီ layer factor ရဲ့ တန်ဖိုးတစ်ခုချင်းစီအတွက် (ဒါမှမဟုတ် control variables နှစ်ခုနဲ့ အထက် ပေါင်းစပ်ထားတဲ့ တန်ဖိုးတွေအတွက်) ဆက်စပ်နေတဲ့ statistics တွေနဲ့ measures တွေကို panel တစ်ခုစီအနေနဲ့ ဖန်တီးပေးမှာ ဖြစ်ပါတယ်။

Example (ဥပမာ)။

အသေးစားကုမ္ပဏီတွေက customers တွေဟာ အကြီးစားကုမ္ပဏီတွေက customers တွေထက် ဝန်ဆောင်မှု (ဥပမာ - သင်တန်းပေးခြင်းနဲ့ အကြံပေးခြင်း) တွေ ဝယ်ယူရာမှာ ပိုပြီး အမြတ်အစွန်းရစေသလား။ Crosstabulation ကနေတစ်ဆင့် သင်သိနိုင်တာကတော့ အသေးစားကုမ္ပဏီ (ဝန်ထမ်း ၅၀၀ အောက်) အများစုက service profits (ဝန်ဆောင်မှုအမြတ်အစွန်း) အများကြီးပေးနိုင်ပြီး၊ အကြီးစားကုမ္ပဏီ (ဝန်ထမ်း ၂၅၀၀ အထက်) အများစုကတော့ service profits နည်းနည်းပဲ ပေးနိုင်တယ် ဆိုတာမျိုးပဲ ဖြစ်ပါတယ်။

Statistics and measures of association (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များနှင့် ဆက်စပ်မှု တိုင်းတာချက်များ

Pearson chi-square, likelihood-ratio chi-square, linear-by-linear association test, Fisher's exact test, Yates' corrected chi-square, Pearson's r, Spearman's rho, contingency coefficient, phi, Cramér's V, symmetric and asymmetric lambdas, Goodman and Kruskal's tau, uncertainty coefficient, gamma, Somers' d, Kendall's tau-b, Kendall's tau-c, eta coefficient, Cohen's kappa, relative risk estimate, odds ratio, McNemar test, Cochran's and Mantel-Haenszel statistics နဲ့ column proportions statistics တို့ ပါဝင်ပါတယ်။

Crosstabs Data Considerations (Crosstabs အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့် အချက်များ)

- **Data (အချက်အလက်)**

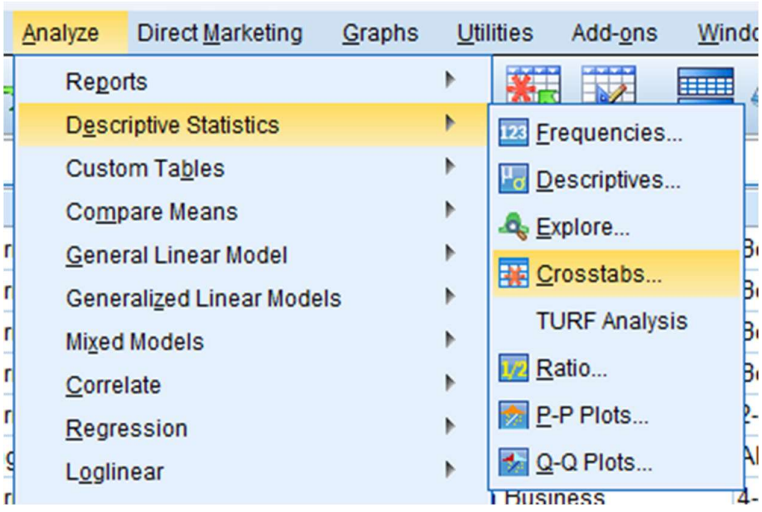
ဇယားထဲက variable တစ်ခုစီရဲ့ categories တွေကို သတ်မှတ်ဖို့အတွက် numeric ဒါမှမဟုတ် string (၈ ဘိုက် သို့မဟုတ် အဲဒီထက်နည်းတဲ့) variable ရဲ့ တန်ဖိုးတွေကို အသုံးပြုပါ။ ဥပမာ- gender အတွက်ဆိုရင် data ကို 1 နဲ့ 2 လို့ဖြစ်စေ၊ male နဲ့ female လို့ဖြစ်စေ code လုပ်နိုင်ပါတယ်။

- **Assumptions (ယူဆချက်များ)**

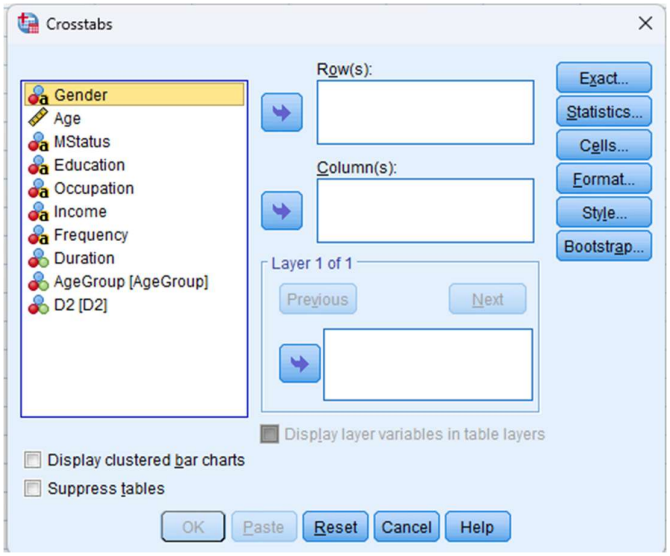
တချို့ statistics တွေနဲ့ measures တွေဟာ ordered categories (ordinal data) သို့မဟုတ် quantitative values (interval ဒါမှမဟုတ် ratio data) တွေဖြစ်တယ်လို့ ယူဆထားပါတယ်။ တခြားအရာတွေကတော့ table variables တွေမှာ အစဉ်လိုက်မရှိတဲ့ categories (nominal data) တွေရှိနေရင် အသုံးပြုနိုင်ပါတယ်။ Chi-square-based statistics (phi, Cramér's V, နဲ့ contingency coefficient) တွေအတွက်ဆိုရင်တော့ data ဟာ multinomial distribution ကနေ ထုတ်ယူထားတဲ့ random sample ဖြစ်သင့်ပါတယ်။ (မှတ်ချက် - Ordinal variables တွေကို numeric codes တွေ ဒါမှမဟုတ် string values တွေအနေနဲ့ ထားနိုင်ပေမယ့်၊ alphabetic အက္ခရာစဉ်အလိုက် စီတာက မှားယွင်းနိုင်တဲ့အတွက် numeric codes တွေက ပိုပြီး ယုံကြည်စိတ်ချရပါတယ်)။

To Obtain Crosstabulations (Crosstabulations ရယူရန် အဆင့်များ)

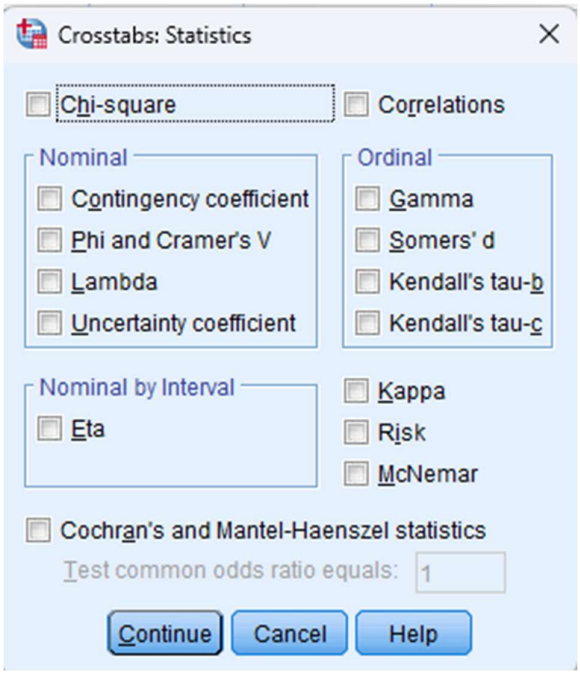
၁။ menus တွေထဲကနေ Analyze > Descriptive Statistics > Crosstabs... ကို ရွေးချယ်ပါ။



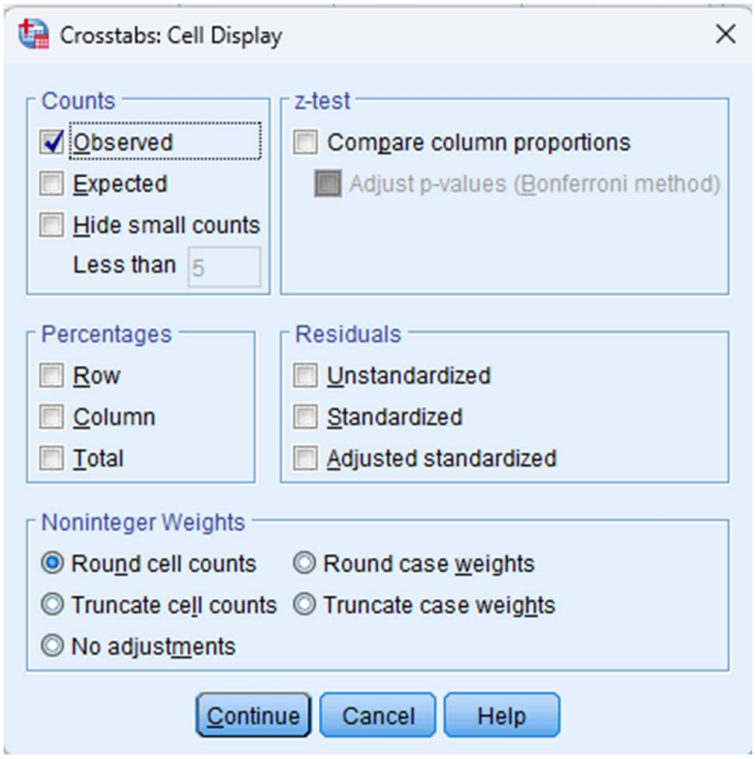
၂။ row variable တစ်ခု သို့မဟုတ် အဲဒီထက်ပိုပြီး ရွေးချယ်ပါ၊ ပြီးရင် column variable တစ်ခု သို့မဟုတ် အဲဒီထက်ပိုပြီး ရွေးချယ်ပါ။ ရွေးချယ်နိုင်သော ထပ်ဆောင်းအချက်များ-



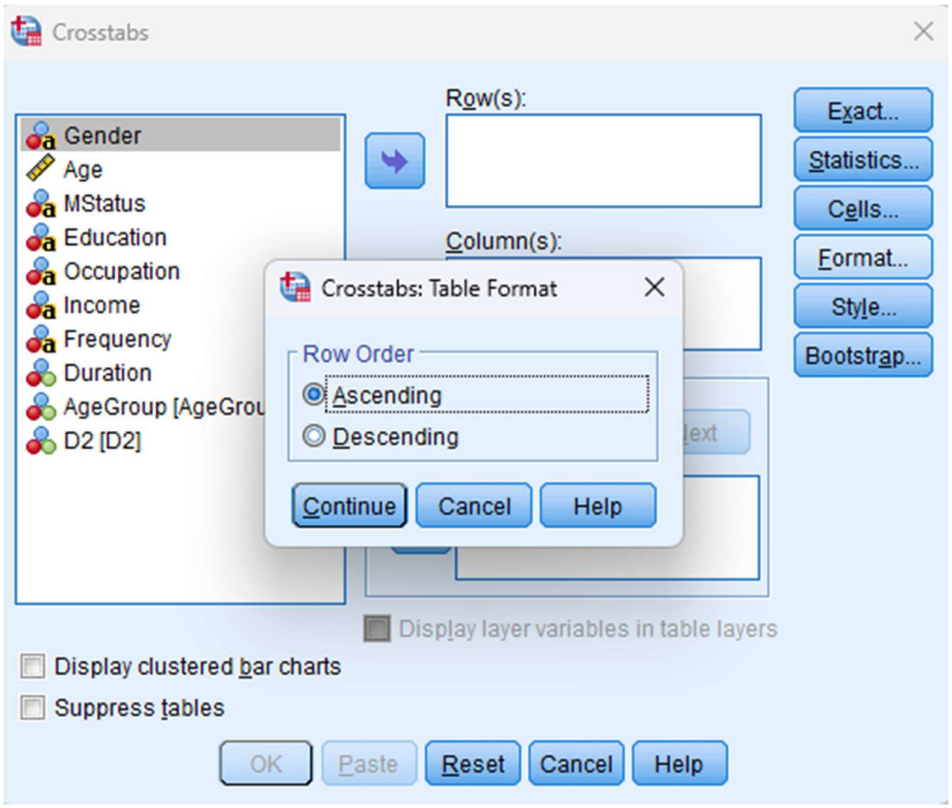
- control variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်နိုင်ပါတယ်။
- two-way tables တွေအတွက် tests တွေနဲ့ measures of association တွေကို လိုချင်ရင် **Statistics** ကို နှိပ်ပါ။



- observed နဲ့ expected values တွေ၊ percentages တွေနဲ့ residuals တွေကို လိုချင်ရင် **Cells** ကို နှိပ်ပါ။



- categories တွေရဲ့ အစီအစဉ်ကို ထိန်းချုပ်ဖို့ **Format** ကို နှိပ်ပါ။



Crosstabs layers (Crosstabs အလွှာများ)

တကယ်လို့ သင်က layer variable (control variable) တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ထားမယ်ဆိုရင်၊ layer variable ရဲ့ category တစ်ခုချင်းစီအတွက် သီးခြား crosstabulation တစ်ခုကို ထုတ်ပေးမှာ ဖြစ်ပါတယ်။ ဥပမာ- row တစ်ခု၊ column တစ်ခုနဲ့ categories နှစ်ခုပါတဲ့ layer variable တစ်ခုရှိရင် two-way table နှစ်ခု ရလာမှာပါ။

Crosstabs clustered bar charts (Crosstabs အုပ်စုလိုက် ဘားချတ်များ)

Cases တွေရဲ့ အုပ်စုတွေအတွက် သင့် data ကို အကျဉ်းချုပ်ပြသဖို့ clustered bar chart ကို ဖန်တီးနိုင်ပါတယ်။ Rows အောက်မှာ သင်သတ်မှတ်ထားတဲ့ variable ရဲ့ တန်ဖိုးတစ်ခုစီအတွက် bar cluster တစ်ခုစီ ရှိနေမှာဖြစ်ပြီး၊ အဲဒီ cluster တစ်ခုစီထဲက bars တွေကို သတ်မှတ်ပေးမှာကတော့ Columns အောက်မှာ သင်ရွေးချယ်ထားတဲ့ variable ပဲ ဖြစ်ပါတယ်။

Crosstabs displaying layer variables in table layers (Crosstabs ဇယားအလွှာများတွင် layer variables များကို ပြသခြင်း)

Layer variables (control variables) တွေကို crosstabulation ဇယားထဲမှာ table layers တွေအနေနဲ့ ပြသဖို့ ရွေးချယ်နိုင်ပါတယ်။ ဒါက row နဲ့ column variables တွေအတွက် overall statistics တွေကို ပြသပေးတဲ့အပြင် layer variables တွေရဲ့ categories တွေကိုပါ drill down (အသေးစိတ် ထပ်မံခွဲခြမ်း) ကြည့်ရှုနိုင်စေပါတယ်။

demo.sav မှာ အောက်ပါအတိုင်း လုပ်ဆောင်ကြည့်ပါ။

၁. Income category in thousands (inccat) ကို Row မှာထားပါ။ Owns PDA (ownpda) ကို column မှာထားပါ။ Level of Education (ed) ကို layer မှာထားပါ။
၂. Display layer variables in table layers ကိုရွေးပါ။
၃. Column ကိုရွေးပါ။
၄. Crosstabs ကို Run ပါ။ , the crosstabulation table ကို နှစ်ချက်နှိပ်ပါ။ Level of education drop down list မှာ College degree ကို ရွေးပါ။.

Crosstabs statistics (Crosstabs ဆိုင်ရာ စာရင်းအင်းများ)

- **Chi-square:** Row နှစ်ခုနဲ့ column နှစ်ခုပါတဲ့ ဇယားတွေအတွက် Pearson chi-square, likelihood-ratio chi-square, Fisher's exact test နဲ့ Yates' corrected chi-square တို့ကို တွက်ချက်ဖို့ ရွေးချယ်ပါ။ Fisher's exact test ကို 5 ထက်ငယ်တဲ့ expected frequency ရှိတဲ့ cell ပါဝင်တဲ့ 2x2 ဇယားတွေအတွက် အသုံးပြုပြီး၊ ကျန်တဲ့ 2x2 ဇယားတွေအတွက်တော့ Yates' corrected chi-square ကို တွက်ချက်ပေးပါတယ်။

- **Correlations:** Row ရော column ပါ ordered values တွေ ပါဝင်နေရင် Spearman's correlation coefficient, rho ကို ရရှိနိုင်ပြီး၊ variables နှစ်ခုလုံးက quantitative ဖြစ်နေရင် Pearson correlation coefficient, r ကို ရရှိနိုင်ပါတယ်။
- **Nominal:** အစဉ်လိုက်မရှိတဲ့ Nominal data တွေအတွက် Contingency coefficient, Phi နဲ့ Cramér's V, Lambda နဲ့ Uncertainty coefficient တို့ကို ရွေးချယ်နိုင်ပါတယ်။
- **Ordinal:** Row ရော column ပါ ordered values တွေ ပါဝင်နေရင် Gamma, Kendall's tau-b, Kendall's tau-c နဲ့ Somers' d တို့ကို ရွေးချယ်နိုင်ပါတယ်။
- **Nominal by Interval:** Variable တစ်ခုက categorical ဖြစ်ပြီး နောက်တစ်ခုက quantitative ဖြစ်နေတဲ့အခါ Eta ကို ရွေးချယ်နိုင်ပါတယ်။
- **Kappa:** Cohen's kappa ဟာ အကဲဖြတ်သူ နှစ်ဦးက တူညီတဲ့အရာတစ်ခုကို အကဲဖြတ်တဲ့အခါ သူတို့ကြားက သဘောတူညီမှု (agreement) ကို တိုင်းတာတာပါ။ 1 ဆိုရင် perfect agreement ကိုပြပြီး 0 ဆိုရင်တော့ chance (တိုက်ဆိုင်မှု) ထက် မပိုဘူးလို့ ဆိုလိုပါတယ်။
- **Risk:** 2x2 ဇယားတွေအတွက် factor တစ်ခု ရှိနေခြင်းနဲ့ အဖြစ်အပျက်တစ်ခု ဖြစ်ပေါ်လာခြင်းကြားက ဆက်စပ်မှုရဲ့ အားကောင်းမှုကို တိုင်းတာပါတယ်။
- **McNemar:** ဆက်စပ်နေတဲ့ dichotomous variables နှစ်ခုအတွက် nonparametric test တစ်ခုပါ။ "before-and-after" (မတိုင်မီနှင့် ပြီးနောက်) ဒီဇိုင်းတွေမှာ အပြောင်းအလဲတွေကို ရှာဖွေဖို့ အသုံးဝင်ပါတယ်။
- **Cochran's and Mantel-Haenszel statistics:** Dichotomous factor variable နဲ့ dichotomous response variable တို့ကြားမှာ အချင်းချင်းမှီခိုမှုမရှိကြောင်း (independence) ကို စမ်းသပ်ဖို့ အသုံးပြုနိုင်ပါတယ်။

Crosstabs cell display (Crosstabs ဆဲလ် ပြသမှု)

- **Counts:** တကယ်တမ်း လေ့လာတွေ့ရှိရတဲ့ (observed) cases အရေအတွက်နဲ့၊ တကယ်လို့ row နဲ့ column variables တွေသာ အချင်းချင်း မမှီခိုဘူးဆိုရင် မျှော်မှန်းရနိုင်တဲ့ (expected) cases အရေအတွက်တွေကို ပြသပေးပါတယ်။ ကိန်းဂဏန်းအသေးလေးတွေကို ဖျောက်ထား (hide) ချင်ရင်လည်း ရပါတယ်။
- **Compare column proportions:** Column proportions တွေကို အချင်းချင်း pairwise နှိုင်းယှဉ်တွက်ချက်ပေးပြီး ဘယ် column စုံတွဲတွေက သိသိသာသာ ကွာခြားမှုရှိလဲဆိုတာကို 0.05 significance level နဲ့ APA-style formatting သုံးပြီး ပြသပေးပါတယ်။ Bonferroni method ကိုသုံးပြီး p-values တွေကို ချိန်ညှိနိုင်ပါတယ်။

- **Percentages:** Row အလိုက်၊ column အလိုက် ဒါမှမဟုတ် ဇယားတစ်ခုလုံးရဲ့ စုစုပေါင်း cases အရေအတွက်အပေါ် မူတည်ပြီး ရာခိုင်နှုန်းတွေကို တွက်ချက်ပေးပါတယ်။
- **Residuals:** Raw unstandardized residuals (observed နဲ့ expected တန်ဖိုးများ ကွာခြားချက်) အပြင် Standardized နဲ့ adjusted standardized residuals တွေကိုပါ ရရှိနိုင်ပါတယ်။
- **Create APA style table:** APA စတိုင် လမ်းညွှန်ချက်တွေနဲ့ ကိုက်ညီတဲ့ output tables တွေကို ဖန်တီးပေးပါတယ်။
- **Noninteger Weights (ကိန်းပြည့်မဟုတ်သော အလေးချိန်များ)** Cell counts တွေဟာ ဆဲလ်တစ်ခုစီမှာပါဝင်တဲ့ cases အရေအတွက်ကို ကိုယ်စားပြုတဲ့အတွက် ပုံမှန်အားဖြင့်တော့ ကိန်းပြည့် (integer) တန်ဖိုးတွေ ဖြစ်ကြပါတယ်။ ဒါပေမယ့် တကယ်လို့ data file က လက်ရှိမှာ fractional values (ဒသမကိန်းများ၊ ဥပမာ - ၁.၂၅) ပါဝင်တဲ့ weight variable တစ်ခုနဲ့ weight ပေးထားမယ်ဆိုရင်တော့၊ cell counts တွေကလည်း fractional values တွေ ဖြစ်လာနိုင်ပါတယ်။ သင်က cell counts တွေကို တွက်ချက်မှု မလုပ်ခင်ဖြစ်စေ၊ လုပ်ပြီးနောက်မှာဖြစ်စေ truncate (အကြွင်းဖြတ်ပစ်ခြင်း) ဒါမှမဟုတ် round (အနီးစပ်ဆုံးကိန်းပြည့်သို့ပြောင်းခြင်း) လုပ်နိုင်သလို၊ ဇယားပြသမှုနဲ့ statistical calculations တွေ နှစ်ခုလုံးအတွက် fractional cell counts ကိုပဲ အသုံးပြုမယ်ဆိုရင်လည်း ရပါတယ်။
- **Round cell counts:** Case weights တွေကို ရှိရင်းစွဲအတိုင်း အသုံးပြုပေမယ့် statistics တွေကို တွက်ချက်မှု မလုပ်ခင်မှာ cells တွေထဲက စုပေါင်းထားတဲ့ weights (accumulated weights) တွေကို အနီးစပ်ဆုံး ကိန်းပြည့်ဖြစ်အောင် ပြောင်းလဲ (round) လိုက်ပါတယ်။
- **Truncate cell counts:** Case weights တွေကို ရှိရင်းစွဲအတိုင်း အသုံးပြုပေမယ့် statistics တွေကို တွက်ချက်မှု မလုပ်ခင်မှာ cells တွေထဲက စုပေါင်းထားတဲ့ weights တွေကို ဒသမပိုင်းဖြတ် (truncate) လိုက်ပါတယ်။
- **Round case weights:** Case weights တွေကို အသုံးမပြုခင်မှာ အနီးစပ်ဆုံး ကိန်းပြည့်ဖြစ်အောင် အရင် ပြောင်းလဲ (round) လိုက်ပါတယ်။
- **Truncate case weights:** Case weights တွေကို အသုံးမပြုခင်မှာ ဒသမပိုင်းကို အရင် ဖြတ် (truncate) လိုက်ပါတယ်။
- **No adjustments:** Case weights တွေကို ရှိရင်းစွဲအတိုင်း အသုံးပြုပြီး fractional cell counts တွေကိုပဲ သုံးပါတယ်။ ဒါပေမယ့် (Sampling နဲ့ Testing တွေမှာပဲ ရနိုင်တဲ့) Exact Statistics ကို တောင်းဆိုထားတဲ့အခါမှာတော့၊ Exact test statistics တွေကို

မတွက်ချက်ခင်မှာ cells တွေထဲက စုပေါင်းထားတဲ့ weights တွေကို truncate ဖြစ်စေ၊ round ဖြစ်စေ လုပ်ပေးပါတယ်။

Crosstabs table format (Crosstabs ဇယား ပုံစံချခြင်း)

row variable ရဲ့ တန်ဖိုးတွေအပေါ် မူတည်ပြီး rows တွေကို အစဉ်လိုက် (ascending) သို့မဟုတ် ပြောင်းပြန်အစဉ်လိုက် (descending) စီစဉ်နိုင်ပါတယ်။

Summarize (အကျဉ်းချုပ်ဖော်ပြခြင်း)

Summarize procedure က grouping variables (အုပ်စုခွဲပြောင်းလဲကိန်း) တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုတဲ့ categories တွေအတွင်းမှာရှိတဲ့ variables တွေအတွက် subgroup statistics (အုပ်စုခွဲစာရင်းအင်းများ) ကို တွက်ချက်ပေးပါတယ်။ Grouping variable ရဲ့ အဆင့် (levels) အားလုံးကို အပြန်အလှန်ယှဉ်တွဲဇယား (crosstabulated) ဖန်တီးပေးပါတယ်။ Statistics တွေကို ဘယ်လိုအစဉ်လိုက် ပြသမလဲဆိုတာကိုလည်း ရွေးချယ်နိုင်ပါတယ်။ Categories အားလုံးတစ်လျှောက် variable တစ်ခုချင်းစီအတွက် summary statistics တွေကိုလည်း ပြသပေးပါတယ်။ Category တစ်ခုချင်းစီမှာရှိတဲ့ data values တွေကို စာရင်းပြုစု (list) တာဖြစ်စေ၊ ဖျောက်ထား (suppress) တာဖြစ်စေ လုပ်နိုင်ပါတယ်။ အကယ်၍ dataset က အရမ်းကြီးနေမယ်ဆိုရင်တော့၊ ပထမဆုံး cases အရေအတွက် n ခုကိုပဲ ပြသဖို့ (ဥပမာ- ပထမဆုံး cases ၁၀၀) ရွေးချယ်နိုင်ပါတယ်။

Example (ဥပမာ)

ဒေသ (region) နဲ့ ဖောက်သည်တွေရဲ့ လုပ်ငန်းအမျိုးအစား (customer industry) အလိုက် ပျမ်းမျှ ထုတ်ကုန်အရောင်းပမာဏ (average product sales amount) က ဘယ်လောက်လဲ။ ဒီကနေတစ်ဆင့် သင်တွေ့ရှိနိုင်တာက အနောက်ဘက်ဒေသ (western region) ရဲ့ ပျမ်းမျှအရောင်းပမာဏက တခြားဒေသတွေထက် အနည်းငယ် ပိုများနေပြီး၊ အနောက်ဘက်ဒေသမှာရှိတဲ့ ကော်ပိုရိတ်ဖောက်သည် (corporate customers) တွေဆီကနေ အများဆုံး ပျမ်းမျှအရောင်းပမာဏကို ရရှိတယ်ဆိုတာမျိုးပဲ ဖြစ်ပါတယ်။

Statistics (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များ)

Sum (ပေါင်းလဒ်), number of cases (cases အရေအတွက်), mean (ပျမ်းမျှကိန်း), median (အလယ်ကိန်း), grouped median, standard error of the mean, minimum, maximum, range, grouping variable ရဲ့ ပထမဆုံး category ရဲ့ variable value, grouping variable ရဲ့ နောက်ဆုံး category ရဲ့ variable value, standard deviation, variance, kurtosis, standard error of kurtosis, skewness, standard error of skewness, percentage of total sum, percentage of total N, percentage of sum in, percentage of N in, geometric mean, နဲ့ harmonic mean တို့ ပါဝင်ပါတယ်။

Summarize Data Considerations (Summarize အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့် အချက်များ)

- **Data (အချက်အလက်)**

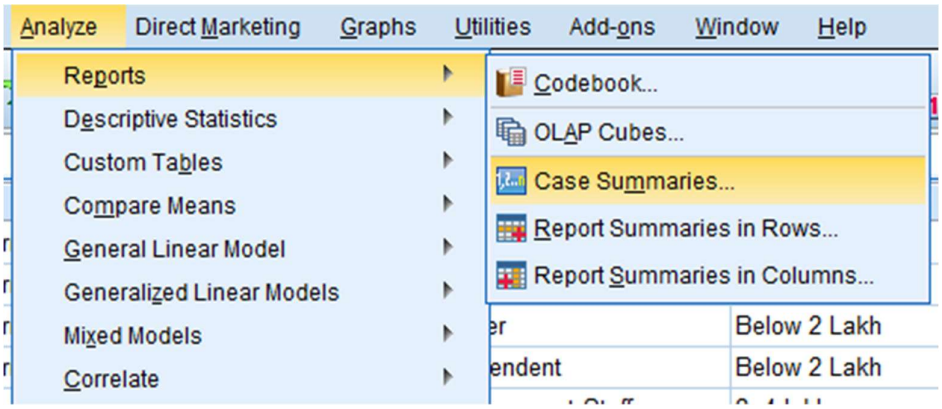
Grouping variables တွေက categorical variables တွေဖြစ်ရမှာဖြစ်ပြီး၊ သူတို့ရဲ့ တန်ဖိုးတွေက ဂဏန်း (numeric) ဖြစ်စေ၊ စာသား (string) ဖြစ်စေ ရပါတယ်။ Categories အရေအတွက်ကတော့ သင့်တင့်မျှတစွာ နည်းပါးသင့်ပါတယ်။ အခြား variables တွေကတော့ အဆင့်ခွဲခြား (rank) နိုင်တဲ့အရာတွေ ဖြစ်ရပါမယ်။

- **Assumptions (ယူဆချက်များ)**

Mean နဲ့ standard deviation လိုမျိုး ရွေးချယ်နိုင်တဲ့ subgroup statistics တချို့ဟာ normal theory အပေါ်မှာ အခြေခံထားတာဖြစ်ပြီး symmetric distributions ရှိတဲ့ quantitative variables တွေအတွက် သင့်လျော်ပါတယ်။ Median နဲ့ range လိုမျိုး robust statistics တွေကတော့ normality assumption နဲ့ ကိုက်ညီတဲ့ သို့မဟုတ် မကိုက်ညီတဲ့ quantitative variables တွေအတွက် အသုံးပြုနိုင်ပါတယ်။

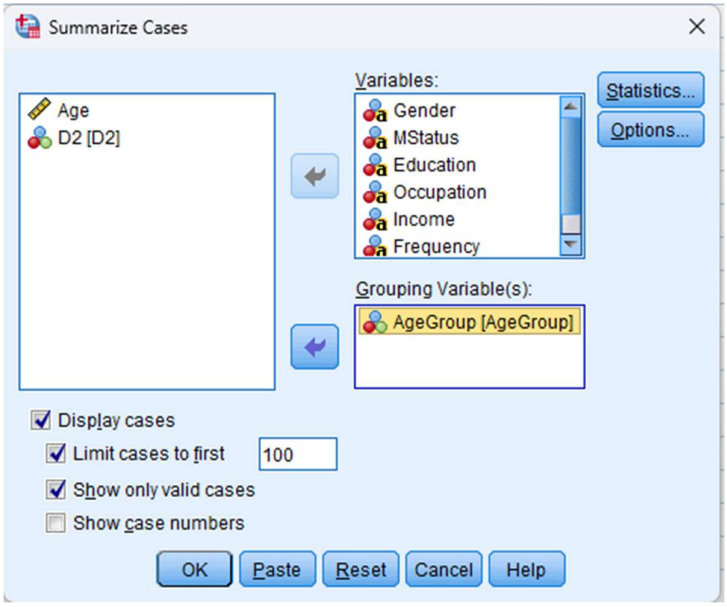
To Obtain Case Summaries (Case Summaries များကို ရယူအသုံးပြုခြင်း)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Reports > Case Summaries...

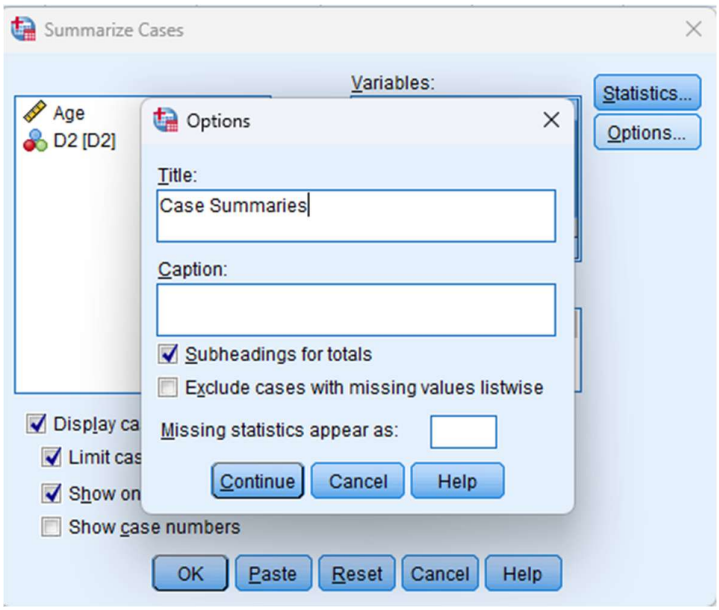


၂။ variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။ ရွေးချယ်နိုင်သော ထပ်ဆောင်းအချက်များ (Optionally, you can:)-

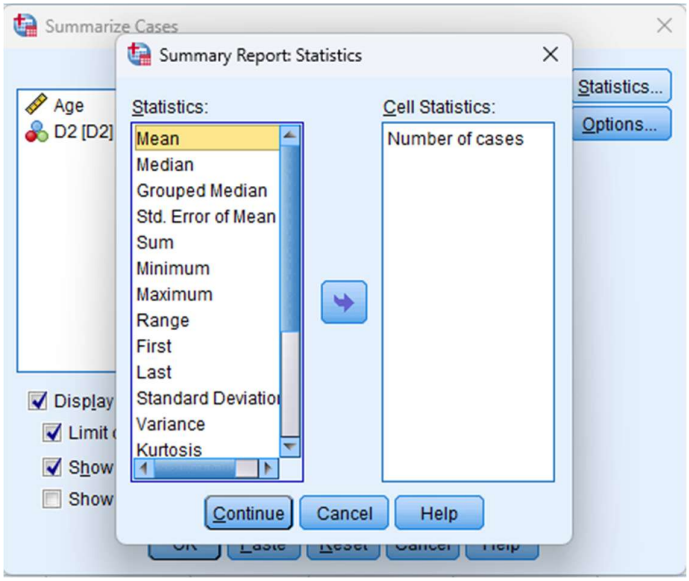
- သင့် data တွေကို subgroups လေးတွေအဖြစ် ခွဲထုတ်ဖို့ grouping variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်နိုင်ပါတယ်။



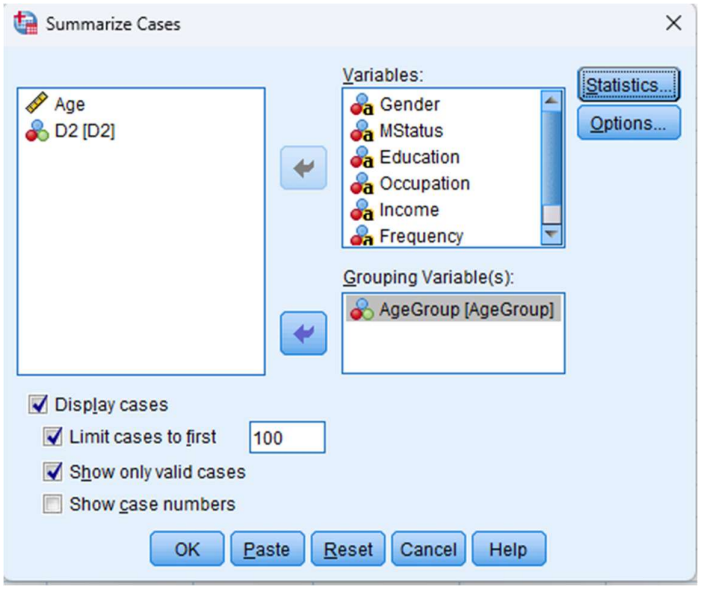
- Output ရဲ့ ခေါင်းစဉ် (title) ကို ပြောင်းလဲဖို့၊ output ဇယားအောက်မှာ စာသား (caption) ထည့်ဖို့၊ ဒါမှမဟုတ် missing values (ပျောက်ဆုံးတန်ဖိုးများ) ရှိတဲ့ cases တွေကို ဖယ်ထုတ်ဖို့ **Options** ကို နှိပ်နိုင်ပါတယ်။



- Optional statistics တွေအတွက် **Statistics** ကို နှိပ်ပါ။



- Subgroup တစ်ခုစီမှာရှိတဲ့ cases တွေကို စာရင်းပြုစုကြည့်ဖို့ **Display cases** ကို ရွေးချယ်ပါ။ ပုံမှန် (By default) အားဖြင့်တော့ စနစ်က သင့်ဖိုင်ထဲမှာရှိတဲ့ ပထမဆုံး cases ၁၀၀ ကိုပဲ ပြသပေးမှာပါ။ **Limit cases to first n** မှာ တန်ဖိုးကို တိုးတာ လျှော့တာ လုပ်နိုင်သလို၊ cases အားလုံးကို ကြည့်ချင်တယ်ဆိုရင်တော့ အဲဒီ option ကို အမှန်ခြစ် ဖြုတ်လိုက်လို့ ရပါတယ်။



Summarize Options (Summarize ဆိုင်ရာ ရွေးချယ်စရာများ)

Summarize က သင့် output ရဲ့ title ကို ပြောင်းလဲဖို့ ဒါမှမဟုတ် output table အောက်မှာ caption ထည့်ဖို့ ခွင့်ပြုပါတယ်။ Titles တွေနဲ့ captions တွေမှာ စာကြောင်းအသစ် (line break) ဆင်းချင်တဲ့ နေရာတိုင်းမှာ \n လို့ ရိုက်ထည့်ပြီး line wrapping ကို ထိန်းချုပ်နိုင်ပါတယ်။ ဒါ့အပြင် totals တွေအတွက် subheadings တွေကို ပြသတာ ဒါမှမဟုတ် ဖျောက်ထားတာမျိုး ရွေးချယ်နိုင်သလို၊ analysis တစ်ခုခုမှာ အသုံးပြုထားတဲ့ variables တွေအတွက် missing values

ရှိနေတဲ့ cases တွေကို ထည့်သွင်းဖို့ ဒါမှမဟုတ် ဖယ်ထုတ်ဖို့လည်း ရွေးချယ်နိုင်ပါတယ်။ Output မှာ missing cases တွေကို အစက်လေး (.) ဒါမှမဟုတ် ခရေပွင့် (*) လေးနဲ့ ပြသလေ့ရှိပါတယ်။ Missing ဖြစ်နေတဲ့အခါ ပေါ်လာစေချင်တဲ့ character, phrase ဒါမှမဟုတ် code တစ်ခုခုကို သင်ကိုယ်တိုင် ရိုက်ထည့်နိုင်ပါတယ်။ အဲဒီလိုမှ မဟုတ်ရင်တော့ output ထဲက missing cases တွေအတွက် ဘာအထူးအစီအစဉ်မှ လုပ်ဆောင်ပေးမှာ မဟုတ်ပါဘူး။

Summarize Statistics (Summarize ဆိုင်ရာ စာရင်းအင်းများ)

Grouping variable တစ်ခုချင်းစီရဲ့ category တစ်ခုစီအတွင်းမှာရှိတဲ့ variables တွေအတွက် အောက်ပါ subgroup statistics တွေထဲက တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်နိုင်ပါတယ်-

- **First / Last.** Data file ထဲမှာ ပထမဆုံး / နောက်ဆုံး တွေ့ရတဲ့ data value ကို ပြသပေးပါတယ်။
- **Geometric / Harmonic Mean.** ဂျီသြမေတြီပျမ်းမျှကိန်း နဲ့ ဟာမိုနစ်ပျမ်းမျှကိန်းတွေ ဖြစ်ပါတယ်။
- **Grouped Median.** Groups တွေအနေနဲ့ code လုပ်ထားတဲ့ data တွေအတွက် တွက်ချက်ထားတဲ့ median ဖြစ်ပါတယ်။
- **Kurtosis / Skewness.** Outliers တွေ ဘယ်လောက်ရှိလဲဆိုတာနဲ့ ဖြန့်ကျက်မှုရဲ့ အချိုးမညီမှု (asymmetry) တွေကို တိုင်းတာတာပါ။
- **Maximum / Minimum / Range.** အကြီးဆုံး၊ အငယ်ဆုံး နဲ့ ၎င်းတို့နှစ်ခုကြား ကွာခြားချက် အကွာအဝေးတို့ ဖြစ်ပါတယ်။
- **Mean / Median / Sum.** ပျမ်းမျှကိန်း၊ အလယ်ကိန်း နဲ့ တန်ဖိုးစုစုပေါင်းတို့ ဖြစ်ပါတယ်။
- **N / Percent of Total N / Percent of Total Sum.** Cases စုစုပေါင်း အရေအတွက်နဲ့ သူတို့ရဲ့ ရာခိုင်နှုန်းတွေ ဖြစ်ပါတယ်။
- **Standard Deviation / Variance.** Mean ကိုဗဟိုပြုပြီး data တွေ ဘယ်လောက် ပြန့်ကျဲနေသလဲဆိုတာ တိုင်းတာမှုတွေပါ။
- **Standard Error (of Kurtosis, Mean, Skewness).** အသီးသီးရဲ့ စံအမှားတွေဖြစ်ပြီး sample တွေကြားက ကွာခြားမှု ဒါမှမဟုတ် normality ကို စစ်ဆေးဖို့ သုံးပါတယ်။

Cell Statistics list ထဲမှာ statistics တွေ ပေါ်နေတဲ့ အစီအစဉ်အတိုင်း output မှာ ပြသပေးသွားမှာ ဖြစ်ပြီး၊ Categories အားလုံးတစ်လျှောက် variable တစ်ခုချင်းစီအတွက် Summary statistics တွေကိုလည်း ပြသပေးမှာ ဖြစ်ပါတယ်။

T Tests (တီ-တက်စ်များ)

IBM SPSS Statistics မှာ အောက်ပါ T tests သုံးမျိုးကို အသုံးပြုနိုင်ပါတယ်-

၁။ **Independent-samples t test (two-sample t test):**

Cases အုပ်စုနှစ်ခုအတွက် variable တစ်ခုရဲ့ means (ပျမ်းမျှကိန်းများ) ကို နှိုင်းယှဉ်ပေးပါတယ်။ အုပ်စုတစ်ခုချင်းစီအတွက် descriptive statistics တွေနဲ့ variances (ကွဲလွဲကိန်း) တူညီမှုရှိမရှိ စစ်ဆေးပေးတဲ့ Levene's test ကို ထုတ်ပေးတဲ့အပြင်၊ equal- နဲ့ unequal-variance t values တွေ၊ means ကွာခြားချက်အတွက် 95% confidence interval တွေကိုပါ ထုတ်ပေးပါတယ်။

၂။ **Paired-samples t test (dependent t test):**

Group တစ်ခုတည်းအတွက် variables နှစ်ခုရဲ့ means ကို နှိုင်းယှဉ်ပေးပါတယ်။ ဒီ test ကို matched pairs ဒါမှမဟုတ် case-control study ဒီဇိုင်းတွေအတွက်လည်း အသုံးပြုပါတယ်။ Output အနေနဲ့ test variables တွေအတွက် descriptive statistics၊ variables တွေကြားက correlation၊ paired differences အတွက် descriptive statistics၊ t test နဲ့ 95% confidence interval တို့ ပါဝင်ပါတယ်။

၃။ **One-sample t test:**

Variable တစ်ခုရဲ့ mean ကို သိထားပြီးသား ဒါမှမဟုတ် မှန်းဆထားတဲ့ (hypothesized) value တစ်ခုနဲ့ နှိုင်းယှဉ်ပေးပါတယ်။ Test variables တွေအတွက် descriptive statistics တွေကို t test နဲ့အတူ ပြသပေးပါတယ်။ Test variable ရဲ့ mean နဲ့ hypothesized test value ကြားက ကွာခြားချက်အတွက် 95% confidence interval ကတော့ default output အနေနဲ့ ပါဝင်လာမှာ ဖြစ်ပါတယ်။

အထက်ပါ Test တစ်ခုချင်းစီရဲ့ အသေးစိတ်ကို အောက်ပါအတိုင်း ဆက်လက်လေ့လာနိုင်ပါတယ်-

1. Independent-Samples T Test

ဒီ procedure က cases အုပ်စုနှစ်ခုအတွက် means ကို နှိုင်းယှဉ်ပေးပြီး၊ t-test effect size တွက်ချက်မှုကိုပါ အလိုအလျောက် လုပ်ဆောင်ပေးပါတယ်။ အကောင်းဆုံးကတော့ subjects တွေကို အုပ်စုနှစ်ခုဆီ random (ကျပန်း) ခွဲဝေပေးသင့်ပါတယ်၊ ဒါမှသာ ရလဒ်ကွာခြားမှုဟာ တခြားအကြောင်းအရင်းတွေကြောင့်မဟုတ်ဘဲ treatment ကြောင့်ဖြစ်တယ်လို့ ဆိုနိုင်မှာပါ။

- **ဥပမာ:** သွေးတိုးလူနာတွေကို ပုံမှန်ဆေး (placebo) သောက်မယ့်အုပ်စုနဲ့ သွေးတိုးကျစေမယ့် အစစ်အမှန်ဆေးသောက်မယ့်အုပ်စု (treatment group) အဖြစ် ကျပန်းခွဲခြားလိုက်တယ် ဆိုပါစို့။ နှစ်လကြာ ကုသပြီးနောက်မှာ အုပ်စုနှစ်ခုရဲ့ ပျမ်းမျှ သွေးပေါင်ချိန်ကို နှိုင်းယှဉ်ဖို့ two-

sample t test ကို သုံးနိုင်ပါတယ်။ လူနာတစ်ယောက်ချင်းစီကို တစ်ကြိမ်စီသာ တိုင်းတာမှာဖြစ်ပြီး သူတို့က အုပ်စုတစ်ခုတည်းမှာပဲ ပါဝင်မှာပါ။

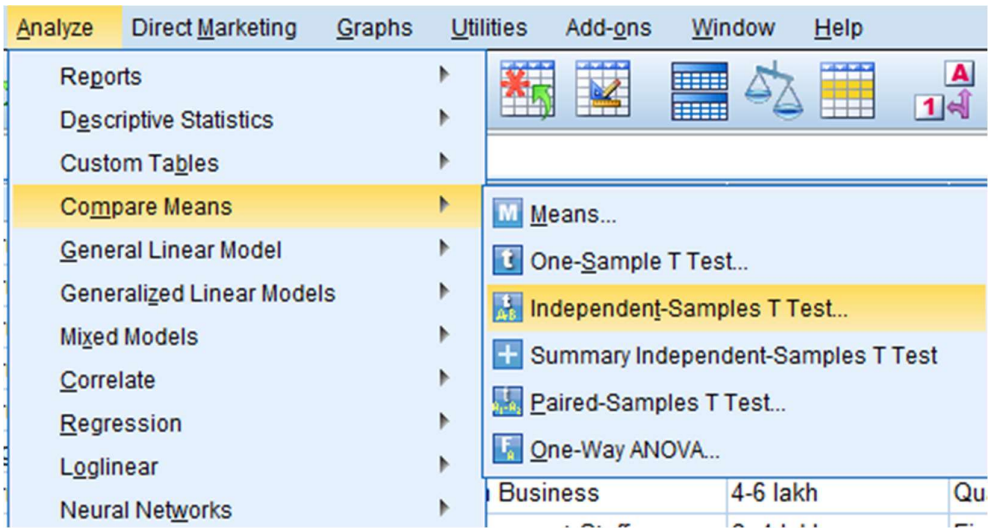
- **Statistics:** Variable တစ်ခုချင်းစီအတွက် sample size, mean, standard deviation, standard error of the mean နဲ့ t-test အတွက် effect size ခန့်မှန်းချက်တွေ ရနိုင်ပါတယ်။ Mean ကွာခြားချက်အတွက်ဆိုရင် mean, standard error နဲ့ confidence interval တို့ ပါဝင်ပြီး၊ Tests အနေနဲ့ Levene's test နဲ့ t tests (pooled-variances နဲ့ separate-variances) တို့ ပါဝင်ပါတယ်။

- **Data Considerations:** Quantitative variable ရဲ့ တန်ဖိုးတွေကို data file ရဲ့ column တစ်ခုတည်းမှာ ထည့်ထားရပါမယ်။ Cases တွေကို အုပ်စုနှစ်ခုခွဲဖို့ တန်ဖိုးနှစ်ခုပါတဲ့ grouping variable ကို သုံးရပါမယ် (ဥပမာ- 1 နဲ့ 2၊ ဒါမှမဟုတ် yes နဲ့ no)။

- **Assumptions:** Equal-variance t test အတွက်ဆိုရင် data တွေဟာ တူညီတဲ့ population variance ရှိပြီး normal distribution ကနေ ကောက်ယူထားတဲ့ independent, random samples တွေ ဖြစ်ရပါမယ်။ Unequal-variance t test အတွက်ကတော့ normal distributions ကနေ ကောက်ယူထားတဲ့ independent, random samples တွေ ဖြစ်ဖို့ပဲ လိုပါတယ်။

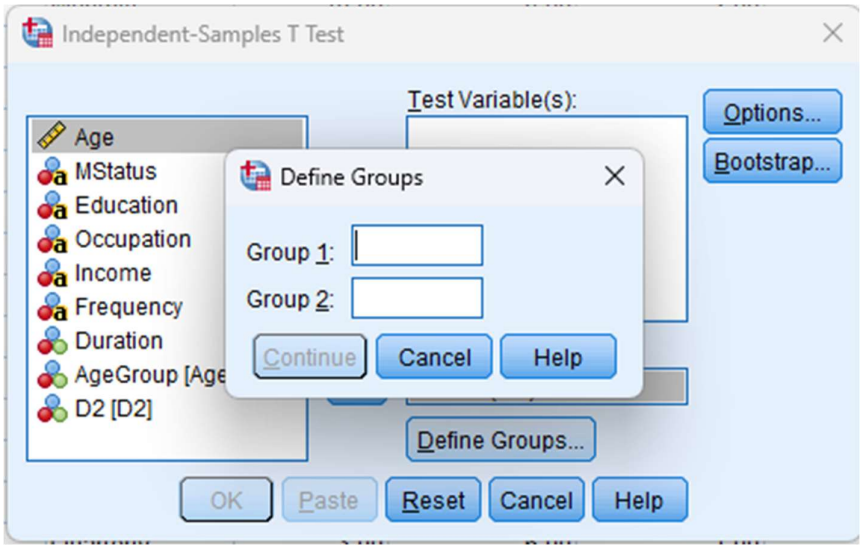
Independent-Samples T Test ရယူအသုံးပြုရန် အဆင့်များ

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Compare Means > Independent-Samples T Test...



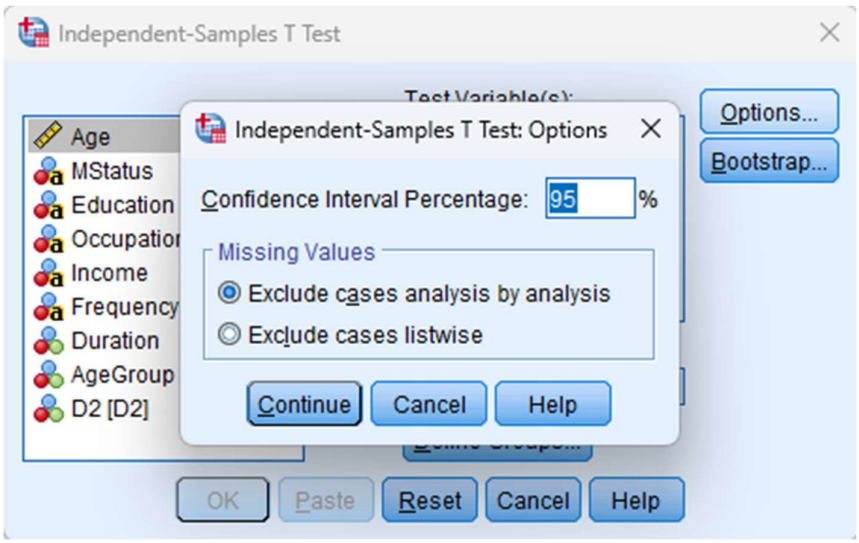
၂။ အရေအတွက်ပြ (quantitative) test variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။ Variable တစ်ခုချင်းစီအတွက် သီးခြား t test တစ်ခုစီကို တွက်ချက်ပေးမှာ ဖြစ်ပါတယ်။

၃။ Grouping variable တစ်ခုတည်းကို ရွေးချယ်ပါ။ ပြီးရင် သင်နှိုင်းယှဉ်ချင်တဲ့ အုပ်စုတွေအတွက် ကုဒ်နှစ်ခု (two codes) ကို သတ်မှတ်ပေးဖို့ **Define Groups** ကို နှိပ်ပါ။



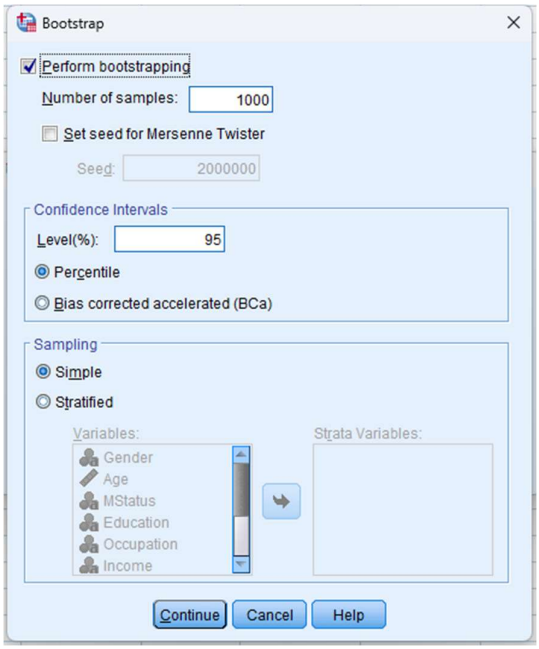
၄။ ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့ သင် အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- t-test effect size တွက်ချက်မှုကို ထိန်းချုပ်ဖို့ **Estimate effect sizes** ကို ရွေးချယ်နိုင်ပါတယ်။
- Missing data တွေကို ဘယ်လိုကိုင်တွယ်မလဲဆိုတာနဲ့ confidence interval level ကို ထိန်းချုပ်ဖို့ **Options** ကို နှိပ်နိုင်ပါတယ်။



- Mean, median, proportion, odds ratio, correlation coefficient ဒါမှမဟုတ် regression coefficient လိုမျိုး ခန့်မှန်းချက်တွေအတွက် ပိုမိုခိုင်မာတဲ့ (robust)

standard errors တွေနဲ့ confidence intervals တွေကို ရယူတွက်ချက်ဖို့ **Bootstrap** ကို နှိပ်နိုင်ပါတယ်။



Independent-Samples T-Test Define Groups (အုပ်စုများ သတ်မှတ်ခြင်း)

ဂဏန်း (numeric) grouping variables တွေအတွက် t-test မှာ သုံးမယ့် အုပ်စုနှစ်ခုကို တန်ဖိုးနှစ်ခု (two values) ဒါမှမဟုတ် cutpoint တစ်ခု သတ်မှတ်ပေးခြင်းဖြင့် ခွဲခြားနိုင်ပါတယ်-

- **Use specified values (သတ်မှတ်ထားသော တန်ဖိုးများကို သုံးရန်)** ။ Group 1 အတွက် တန်ဖိုးတစ်ခု ထည့်ပါ။ ပြီးတော့ Group 2 အတွက် နောက်တန်ဖိုးတစ်ခု ထည့်ပါ။ အခြားတန်ဖိုးတွေပါတဲ့ cases တွေကိုတော့ analysis ထဲကနေ ဖယ်ထုတ်ထားပါလိမ့်မယ်။ ဂဏန်းတွေက ကိန်းပြည့် (integers) ဖြစ်နေစရာ မလိုပါဘူး (ဥပမာ- 6.25 နဲ့ 12.5 တို့လို ဒသမကိန်းတွေလည်း သုံးလို့ရပါတယ်) ။
- **Cutpoint (ဖြတ်မှတ်)**။ Grouping variable ရဲ့ တန်ဖိုးတွေကို အုပ်စုနှစ်ခုအဖြစ် ခွဲထုတ်ပေးမယ့် ဂဏန်းတစ်ခုကို ရိုက်ထည့်ပါ။ Cutpoint ထက်ငယ်တဲ့ တန်ဖိုးရှိတဲ့ cases အားလုံးက အုပ်စုတစ်စုဖြစ်သွားမှာဖြစ်ပြီး၊ Cutpoint နဲ့ ညီမျှတဲ့ ဒါမှမဟုတ် ကြီးတဲ့ cases တွေကတော့ နောက်အုပ်စုတစ်စု ဖြစ်သွားမှာပါ။

စာသား (string) grouping variables တွေအတွက်ကတော့ Group 1 အတွက် စာသားတစ်ခု ထည့်ပြီး၊ Group 2 အတွက် နောက်စာသားတစ်ခု (ဥပမာ- yes နဲ့ no လိုမျိုး) ထည့်ပေးရပါမယ်။ အခြားစာသားတွေပါတဲ့ cases တွေကိုတော့ analysis ထဲကနေ ဖယ်ထုတ်ထားမှာ ဖြစ်ပါတယ်။

Independent-Samples T Test Options (ရွေးချယ်စရာများ)

Confidence Interval (ယုံကြည်မှု အတိုင်းအတာ) ။

ပုံမှန်အားဖြင့် (By default) means ကွာခြားချက်အတွက် 95% confidence interval ကို ပြသပေးပါတယ်။ တခြား confidence level ကို လိုချင်တယ်ဆိုရင်တော့ 1 ကနေ 99 ကြား တန်ဖိုးတစ်ခုခုကို ရိုက်ထည့်နိုင်ပါတယ်။

Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)။

သင်က variables အများကြီးကို test လုပ်တဲ့အခါ၊ variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး data တွေ missing ဖြစ်နေ (ပျောက်ဆုံးနေ) မယ်ဆိုရင်၊ ဘယ် cases တွေကို ထည့်သွင်းရမယ် (ဒါမှမဟုတ် ဖယ်ထုတ်ရမယ်) ဆိုတာကို procedure ကို ညွှန်ကြားနိုင်ပါတယ်-

- **Exclude cases analysis by analysis။** t test တစ်ခုစီဟာ သူ test လုပ်မယ့် variables တွေအတွက် valid ဖြစ်တဲ့ (မှန်ကန်ပြည့်စုံတဲ့) data ရှိတဲ့ cases အားလုံးကို အသုံးပြုပါတယ်။ ဒါကြောင့် test တစ်ခုနဲ့တစ်ခုကြားမှာ sample sizes တွေ ကွာခြားနိုင်ပါတယ်။
- **Exclude cases listwise။** t test တစ်ခုစီဟာ တောင်းဆိုထားတဲ့ t tests တွေအားလုံးမှာ အသုံးပြုမယ့် variables အားလုံးအတွက် valid ဖြစ်တဲ့ data ရှိတဲ့ cases တွေကိုသာ အသုံးပြုပါတယ်။ ဒါကြောင့် test အားလုံးအတွက် sample size က ပုံသေ (constant) ဖြစ်နေမှာပါ။

2. Paired-Samples T Test

ဒီ procedure က group တစ်ခုတည်းအတွက် variables နှစ်ခုရဲ့ means ကို နှိုင်းယှဉ်ပေးပါတယ်။ Case တစ်ခုချင်းစီအတွက် variables နှစ်ခုရဲ့ values ကွာခြားချက်တွေကို တွက်ချက်ပေးပြီး အဲဒီပျမ်းမျှကွာခြားချက်က 0 နဲ့ ကွာခြားမှု ရှိမရှိ စမ်းသပ်ပေးပါတယ်။

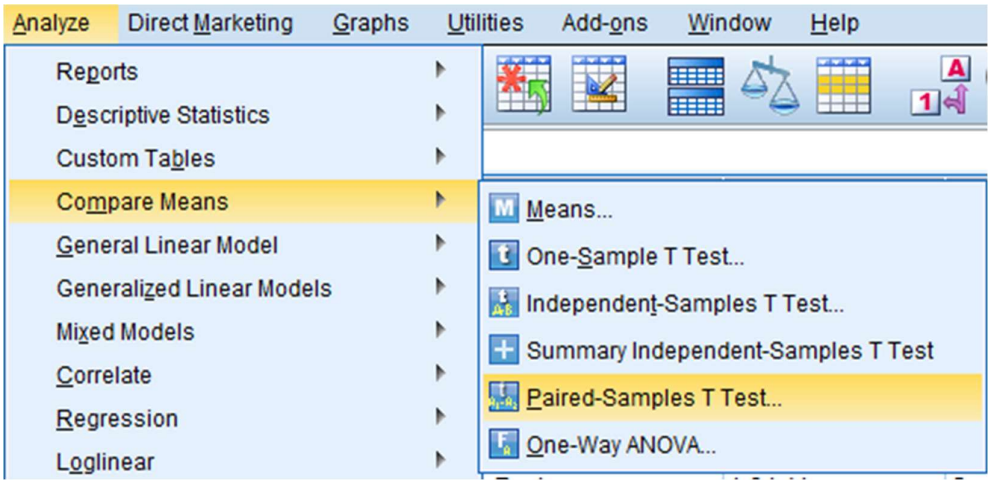
- **ဥပမာ:** သွေးတိုးရောဂါ လေ့လာမှုတစ်ခုမှာ လူနာအားလုံးကို လေ့လာမှုအစမှာ တိုင်းတာတယ်၊ ကုသမှုပေးတယ်၊ ပြီးတော့ ထပ်တိုင်းတာတယ် ဆိုပါစို့။ ဒါဆိုရင် လူနာတစ်ယောက်စီမှာ before နဲ့ after ဆိုတဲ့ တိုင်းတာချက် နှစ်ခုစီ ရှိနေမှာ ဖြစ်ပါတယ်။
- **Statistics:** Variable တစ်ခုချင်းစီအတွက် mean, sample size, standard deviation, နဲ့ standard error of the mean ပါဝင်ပါတယ်။ Variable pairs တွေအတွက် correlation, average difference in means, t test, confidence interval နဲ့ effect size ခန့်မှန်းချက်တို့ ရနိုင်ပါတယ်။
- **Data Considerations:** Paired test တစ်ခုစီအတွက် quantitative variables နှစ်ခု သတ်မှတ်ပေးဖို့ လိုအပ်ပါတယ်။

- **Assumptions:** Pair တစ်ခုချင်းစီအတွက် တိုင်းတာမှုတွေကို တူညီတဲ့ အခြေအနေအောက်မှာ ပြုလုပ်ထားသင့်ပါတယ်။ Mean differences တွေက normally distributed ဖြစ်သင့်ပါတယ်။

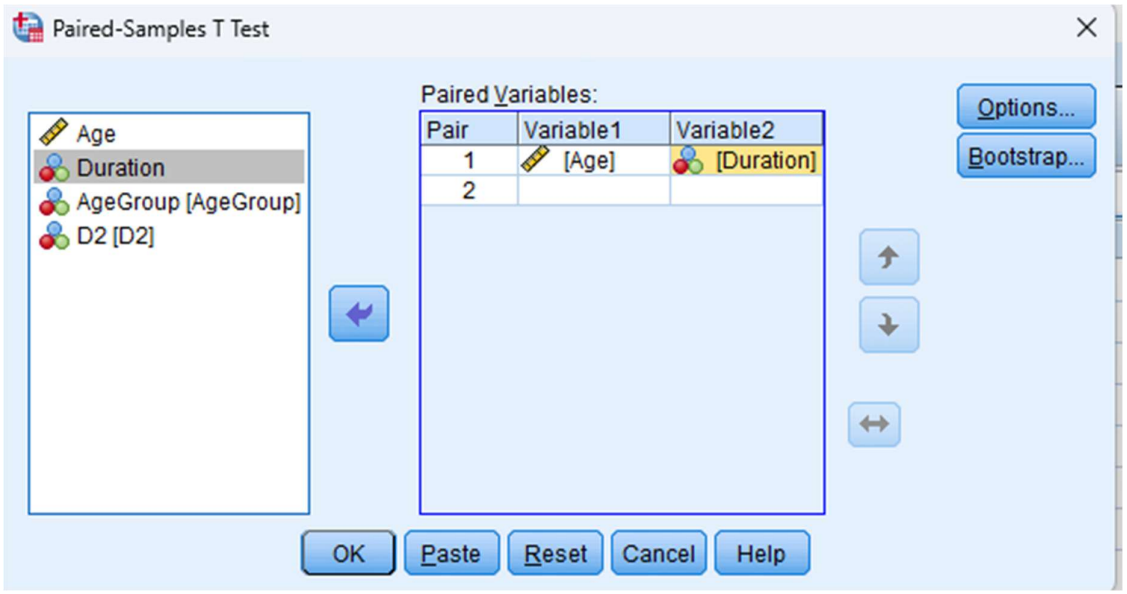
Obtaining a Paired-Samples T Test (Paired-Samples T Test ရယူအသုံးပြုရန် အဆင့်များ)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ-

Analyze > Compare Means > Paired-Samples T Test...



၂။ variable စုံတွဲ (pairs of variables) တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။



၃။ ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်အနေနဲ့ (Optionally)၊ **Estimate effect sizes** (effect size ခန့်မှန်းခြင်း) ရွေးချယ်စရာတစ်ခုကို ပြောင်းလဲ/ရွေးချယ်နိုင်ပါတယ်။ ဒီ settings တွေက

variable စုံတွဲတစ်ခုစီအတွက် Cohen's d နဲ့ Hedges' correction တွေကို ခန့်မှန်းတွက်ချက်ရာမှာ standardizer ကို ဘယ်လို တွက်ချက်မလဲဆိုတာကို ထိန်းချုပ်ပေးပါတယ်။

- **Standard deviation of the difference (ကွာခြားချက်ရဲ့ စံသွေဖည်ကိန်း) -** Effect size ကို ခန့်မှန်းတွက်ချက်ရာမှာ သုံးတဲ့ ပိုင်းခြေ (denominator) ဖြစ်ပါတယ်။ Cohen's d က ပျမ်းမျှကွာခြားချက် (mean difference) ရဲ့ sample standard deviation ကို အသုံးပြုပါတယ်။ Hedges' correction ကတော့ correction factor တစ်ခုနဲ့ ချိန်ညှိ (adjust) ထားတဲ့ ပျမ်းမျှကွာခြားချက်ရဲ့ sample standard deviation ကို အသုံးပြုပါတယ်။

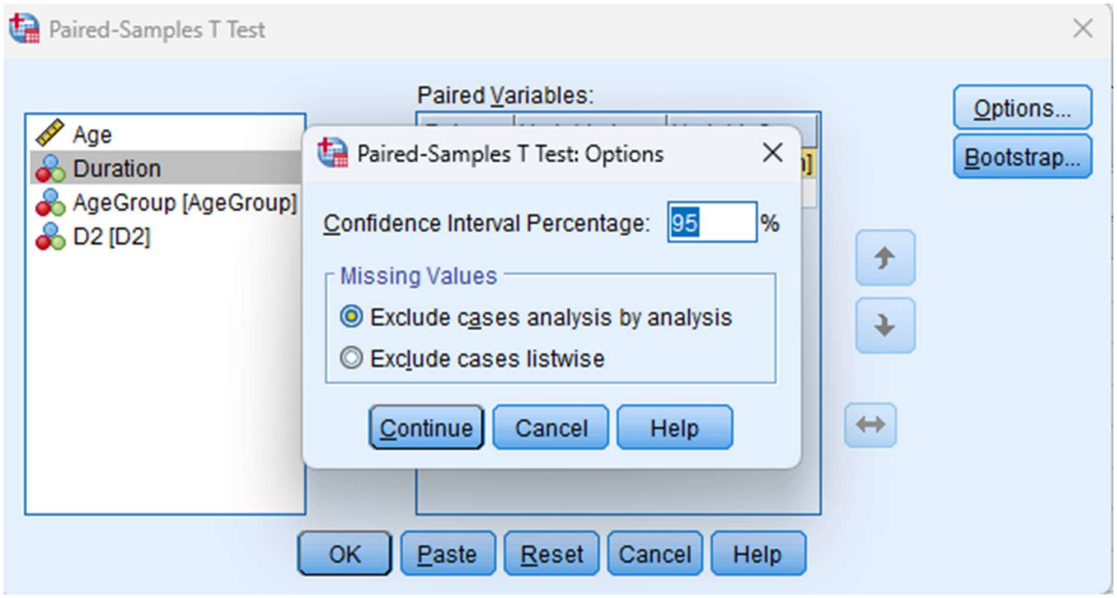
- **Corrected standard deviation of the difference (ချိန်ညှိထားသော ကွာခြားချက်ရဲ့ စံသွေဖည်ကိန်း) -** Effect size ကို ခန့်မှန်းတွက်ချက်ရာမှာ သုံးတဲ့ ပိုင်းခြေ ဖြစ်ပါတယ်။ Cohen's d က တိုင်းတာချက်တွေကြားက correlation နဲ့ ချိန်ညှိထားတဲ့ ပျမ်းမျှကွာခြားချက်ရဲ့ sample standard deviation ကို အသုံးပြုပါတယ်။ Hedges' correction ကတော့ တိုင်းတာချက်တွေကြားက correlation နဲ့ ချိန်ညှိထားတဲ့အပြင် correction factor တစ်ခုပါ ပေါင်းထည့်ထားတဲ့ ပျမ်းမျှကွာခြားချက်ရဲ့ sample standard deviation ကို အသုံးပြုပါတယ်။

- **Average of variances (Variances များရဲ့ ပျမ်းမျှ) -** Effect size ကို ခန့်မှန်းတွက်ချက်ရာမှာ သုံးတဲ့ ပိုင်းခြေ ဖြစ်ပါတယ်။ Cohen's d က တိုင်းတာချက်တွေရဲ့ ပျမ်းမျှ variance (average variance of measures) ကို နှစ်ထပ်ကိန်းရင်း (square root) ရှာပြီး အသုံးပြုပါတယ်။ Hedges' correction ကတော့ တိုင်းတာချက်တွေရဲ့ ပျမ်းမျှ variance ကို နှစ်ထပ်ကိန်းရင်း ရှာထားတဲ့အပြင် correction factor တစ်ခုပါ ပေါင်းထည့်ပြီး အသုံးပြုပါတယ်။

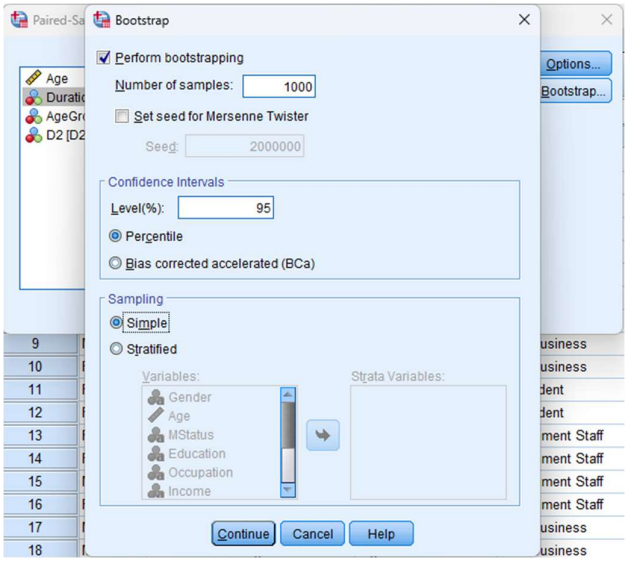
၄။ ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့ သင် အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- t-test effect size တွက်ချက်မှုကို ထိန်းချုပ်ဖို့ **Estimate effect sizes** ကို ရွေးချယ်နိုင်ပါတယ်။ အဲဒီ setting ကို ရွေးချယ်လိုက်ရင်၊ variable စုံတွဲတစ်ခုစီအတွက် Cohen's d နဲ့ Hedges' correction တွေကို ခန့်မှန်းတွက်ချက်ရာမှာ standardizer ကို ဘယ်လို တွက်ချက်မလဲဆိုတာကို ထပ်မံ ထိန်းချုပ်နိုင်ပါတယ်။

- Missing data တွေကို ဘယ်လိုကိုင်တွယ်မလဲဆိုတာနဲ့ confidence interval level ကို ထိန်းချုပ်ဖို့ **Options** ကို နှိပ်နိုင်ပါတယ်။



- Mean, median, proportion, odds ratio, correlation coefficient ဒါမှမဟုတ် regression coefficient လိုမျိုး ခန့်မှန်းချက်တွေအတွက် ပိုမိုခိုင်မာတဲ့ (robust) standard errors တွေနဲ့ confidence intervals တွေကို ရယူတွက်ချက်ဖို့ **Bootstrap** ကို နှိပ်နိုင်ပါတယ်။



Paired-Samples T Test Options (ရွေးချယ်စရာများ)

Confidence Interval (ယုံကြည်မှု အတိုင်းအတာ)

ပုံမှန်အားဖြင့် (By default) means ကွာခြားချက်အတွက် 95% confidence interval ကို ပြသပေးပါတယ်။ တခြား confidence level ကို လိုချင်တယ်ဆိုရင်တော့ 1 ကနေ 99 ကြား တန်ဖိုးတစ်ခုခုကို ရိုက်ထည့်နိုင်ပါတယ်။

Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)

သင်က variables အများကြီးကို test လုပ်တဲ့အခါ၊ variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး data တွေ missing ဖြစ်နေ (ပျောက်ဆုံးနေ) မယ်ဆိုရင်၊ ဘယ် cases တွေကို ထည့်သွင်းရမယ် (ဒါမှမဟုတ် ဖယ်ထုတ်ရမယ်) ဆိုတာကို procedure ကို ညွှန်ကြားနိုင်ပါတယ်-

- **Exclude cases analysis by analysis။** t test တစ်ခုစီဟာ သူ test လုပ်မယ့် variable စုံတွဲအတွက် valid ဖြစ်တဲ့ (မှန်ကန်ပြည့်စုံတဲ့) data ရှိတဲ့ cases အားလုံးကို အသုံးပြုပါတယ်။ ဒါကြောင့် test တစ်ခုနဲ့တစ်ခုကြားမှာ sample sizes တွေ ကွာခြားနိုင်ပါတယ်။
- **Exclude cases listwise။** t test တစ်ခုစီဟာ တောင်းဆိုထားတဲ့ test လုပ်မယ့် variable စုံတွဲတွေ အားလုံးအတွက် valid ဖြစ်တဲ့ data ရှိတဲ့ cases တွေကိုသာ အသုံးပြုပါတယ်။ ဒါကြောင့် test အားလုံးအတွက် sample size က ပုံသေ (constant) ဖြစ်နေမှာပါ။

T TEST Command Additional Features (T TEST Command ၏ အခြားလုပ်ဆောင်ချက်များ)

Command syntax language ကိုအသုံးပြုပြီး အောက်ပါတို့ကိုလည်း လုပ်ဆောင်နိုင်ပါသေးတယ်-

- Command တစ်ခုတည်းကို run ရုံနဲ့ one-sample ရော independent-samples t tests တွေကိုပါ ထုတ်ယူနိုင်ပါတယ်။
- Paired t test တစ်ခုမှာ variable တစ်ခုကို စာရင်းထဲမှာရှိတဲ့ variable တစ်ခုချင်းစီနဲ့ ယှဉ်ပြီး test လုပ်နိုင်ပါတယ် (PAIRS subcommand ကို သုံးပြီး) ။
- t-test effect size ခန့်မှန်းတွက်ချက်မှုကို ထိန်းချုပ်နိုင်ပါတယ် (ES subcommand ကို သုံးပြီး) ။ (ပြီးပြည့်စုံတဲ့ syntax အချက်အလက်တွေအတွက် Command Syntax Reference ကို ကြည့်ရှုနိုင်ပါတယ်။)

3. One-Sample T Test

ဒီ procedure က variable တစ်ခုတည်းရဲ့ mean ဟာ သတ်မှတ်ထားတဲ့ ကိန်းသေတစ်ခု (specified constant / hypothesized value) နဲ့ ကွာခြားမှု ရှိမရှိကို စမ်းသပ်ပေးပါတယ်။

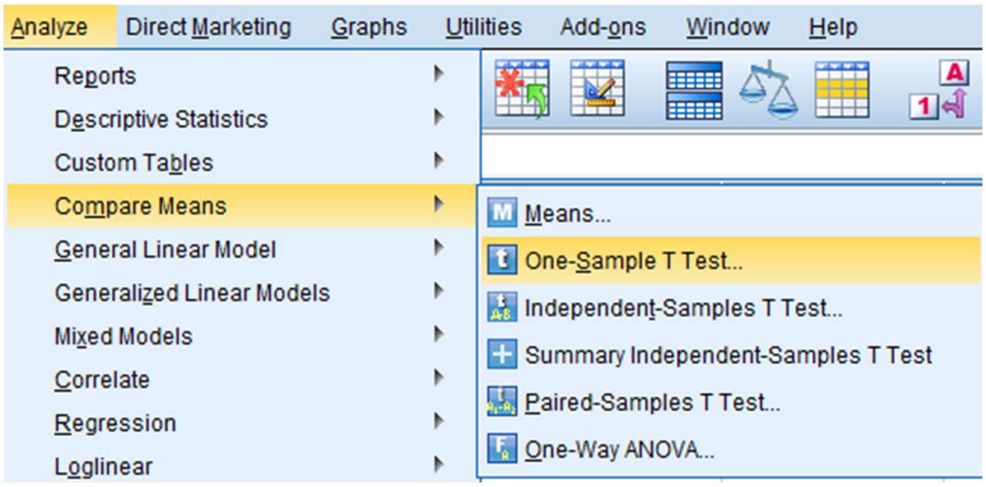
- **ဥပမာ:** သုတေသီတစ်ယောက်က ကျောင်းသားတစ်စုရဲ့ ပျမ်းမျှ IQ score ဟာ 100 နဲ့ ကွာခြားမှု ရှိမရှိ စမ်းသပ်ချင်တာမျိုး ဖြစ်ပါတယ်။
- **Statistics:** Test variable တစ်ခုစီအတွက် mean, standard deviation, standard error of the mean နဲ့ effect size ခန့်မှန်းချက်။ Data value တစ်ခုချင်းစီနဲ့ hypothesized test value

ကြားက ပျမ်းမျှကွာခြားချက် (average difference) ၊ အဲဒီကွာခြားချက်က 0 ဖြစ်မဖြစ် စမ်းသပ်တဲ့ t test နဲ့ confidence interval တို့ ပါဝင်ပါတယ်။

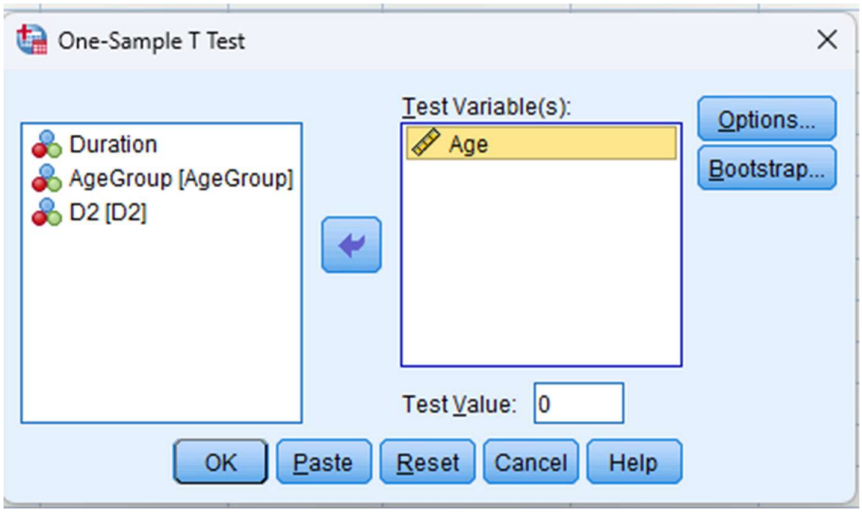
- **Data Considerations:** Quantitative variable တစ်ခုကိုရွေးပြီး ယှဉ်တွဲစစ်ဆေးမယ့် hypothesized test value တစ်ခုကို ရိုက်ထည့်ပေးရပါမယ်။
- **Assumptions:** Data တွေက normally distributed ဖြစ်တယ်လို့ ယူဆထားပေမယ့်၊ normality ကနေ သွေဖည်နေရင်တောင် ဒီ test က တော်တော်လေး robust ဖြစ် (အမှားအယွင်းခံနိုင်) ပါတယ်။

Obtaining a One-Sample T Test (One-Sample T Test ရယူအသုံးပြုရန် အဆင့်များ)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Compare Means > One-Sample T Test...



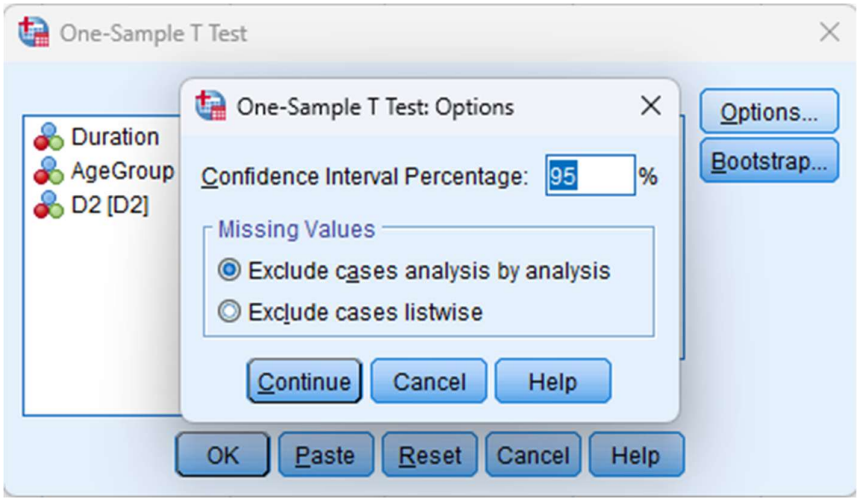
၂။ တူညီတဲ့ မှန်းဆထားတဲ့ တန်ဖိုး (hypothesized value) တစ်ခုနဲ့ ယှဉ်ပြီး test လုပ်မယ့် variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပေးပါ။



၃။ sample mean တစ်ခုစီကို ဘယ်တန်ဖိုးနဲ့ ယှဉ်ပြီး နှိုင်းယှဉ်မလဲဆိုတာအတွက် ဂဏန်း (numeric) test value တစ်ခုကို ရိုက်ထည့်ပါ။

၄။ ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့ သင် အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- t-test effect size ခန့်မှန်းတွက်ချက်မှုကို ထိန်းချုပ်ဖို့ **Estimate effect sizes** ကို ရွေးချယ်နိုင်ပါတယ်။
- Missing data တွေကို ဘယ်လိုကိုင်တွယ်မလဲဆိုတာနဲ့ confidence interval level ကို ထိန်းချုပ်ဖို့ **Options** ကို နှိပ်နိုင်ပါတယ်။



One-Sample T Test Options (ရွေးချယ်စရာများ)

Confidence Interval (ယုံကြည်မှု အတိုင်းအတာ)

ပုံမှန်အားဖြင့် (By default) mean နဲ့ hypothesized test value တို့ကြားက ကွာခြားချက်အတွက် 95% confidence interval ကို ပြသပေးပါတယ်။ တခြား confidence level ကို လိုချင်တယ်ဆိုရင်တော့ 1 ကနေ 99 ကြား တန်ဖိုးတစ်ခုခုကို ရိုက်ထည့်နိုင်ပါတယ်။

Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)

သင်က variables အများကြီးကို test လုပ်တဲ့အခါ၊ variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး data တွေ missing ဖြစ်နေ (ပျောက်ဆုံးနေ) မယ်ဆိုရင်၊ ဘယ် cases တွေကို ထည့်သွင်းရမယ် (ဒါမှမဟုတ် ဖယ်ထုတ်ရမယ်) ဆိုတာကို procedure ကို ညွှန်ကြားနိုင်ပါတယ်-

- **Exclude cases analysis by analysis**။ t test တစ်ခုစီဟာ သူ test လုပ်မယ့် variable အတွက် valid ဖြစ်တဲ့ (မှန်ကန်ပြည့်စုံတဲ့) data ရှိတဲ့ cases အားလုံးကို အသုံးပြုပါတယ်။ ဒါကြောင့် test တစ်ခုနဲ့တစ်ခုကြားမှာ sample sizes တွေ ကွာခြားနိုင်ပါတယ်။
- **Exclude cases listwise**။ t test တစ်ခုစီဟာ တောင်းဆိုထားတဲ့ t tests တွေထဲက တစ်ခုခုမှာ အသုံးပြုမယ့် variables အားလုံးအတွက် valid ဖြစ်တဲ့ data ရှိတဲ့ cases တွေကိုသာ အသုံးပြုပါတယ်။ ဒါကြောင့် test အားလုံးအတွက် sample size က ပုံသေ (constant) ဖြစ်နေမှာပါ။

T TEST Command Additional Features (T TEST Command ၏ အခြားလုပ်ဆောင်ချက်များ)

Command syntax language ကိုအသုံးပြုပြီး အောက်ပါတို့ကိုလည်း လုပ်ဆောင်နိုင်ပါသေးတယ်-

- Command တစ်ခုတည်းကို run ရုံနဲ့ one-sample ရော independent-samples t tests တွေကိုပါ ထုတ်ယူနိုင်ပါတယ်။
- Paired t test တစ်ခုမှာ variable တစ်ခုကို စာရင်းထဲမှာရှိတဲ့ variable တစ်ခုချင်းစီနဲ့ ယှဉ်ပြီး test လုပ်နိုင်ပါတယ် (PAIRS subcommand ကို သုံးပြီး) ။
- t-test effect size ခန့်မှန်းတွက်ချက်မှုကို ထိန်းချုပ်နိုင်ပါတယ် (ES subcommand ကို သုံးပြီး) ။

(ပြီးပြည့်စုံတဲ့ syntax အချက်အလက်တွေအတွက် Command Syntax Reference ကို ကြည့်ရှုနိုင်ပါတယ်။)

One-Way ANOVA (တစ်ဖက်သတ် ကွဲလွဲကိန်း ခွဲခြမ်းစိတ်ဖြာခြင်း)

One-Way ANOVA procedure ဟာ အရေအတွက်ပြ (quantitative) dependent variable တစ်ခုကို factor (independent) variable တစ်ခုတည်းနဲ့ ခွဲခြားပြီး one-way analysis of variance ကို ထုတ်ပေးတဲ့အပြင်၊ one-way ANOVA ရဲ့ effect size ကိုလည်း ခန့်မှန်းတွက်ချက်ပေးပါတယ်။ Analysis of variance (ANOVA) ဆိုတာ အုပ်စုအများကြီးရဲ့ ပျမ်းမျှကိန်း (means) တွေ တူညီမှု ရှိမရှိ ဆိုတဲ့ hypothesis ကို စမ်းသပ်ဖို့ သုံးတာပါ။ ဒီနည်းလမ်းက two-sample t test ကို ပိုမိုကျယ်ပြန့်အောင် ထပ်မံချဲ့ထွင်ထားတာပဲ ဖြစ်ပါတယ်။

Means တွေကြားမှာ ကွာခြားချက်တွေ ရှိနေတယ်လို့ ဆုံးဖြတ်ရုံသာမကဘဲ၊ ဘယ် means တွေက ကွာခြားနေတာလဲ ဆိုတာကိုပါ သင် သိချင်ပါလိမ့်မယ်။ Means တွေကို နှိုင်းယှဉ်ဖို့အတွက် test နှစ်မျိုးရှိပါတယ်- a priori contrasts နဲ့ post hoc tests တို့ ဖြစ်ပါတယ်။ Contrasts ဆိုတာ စမ်းသပ်မှု (experiment) မလုပ်ခင်ကတည်းက ကြိုတင်သတ်မှတ်ထားတဲ့ tests တွေဖြစ်ပြီး၊ post hoc tests တွေကတော့ စမ်းသပ်မှုလုပ်ပြီးသွားမှ လုပ်ဆောင်တဲ့ tests တွေပါ။ Categories တွေတစ်လျှောက် ပြောင်းလဲသွားတဲ့ trends တွေကိုလည်း သင် စမ်းသပ်နိုင်ပါသေးတယ်။

Example (ဥပမာ)

ဒိုးနတ် (Doughnuts) တွေကို ကြော်တဲ့အခါ အဆီစုပ်ယူမှုပမာဏ ကွာခြားတတ်ပါတယ်။ မြေပဲဆီ၊ ပြောင်းဖူးဆီ နဲ့ ဝက်ဆီ ဆိုတဲ့ အဆီသုံးမျိုးကို သုံးပြီး စမ်းသပ်မှုတစ်ခု လုပ်ကြည့်မယ် ဆိုပါစို့။ မြေပဲဆီနဲ့ ပြောင်းဖူးဆီက unsaturated fats (မပြည့်ဝဆီ) တွေဖြစ်ပြီး၊ ဝက်ဆီကတော့ saturated fat (ပြည့်ဝဆီ) ဖြစ်ပါတယ်။ အဆီစုပ်ယူမှု ပမာဏဟာ အသုံးပြုလိုက်တဲ့ အဆီအမျိုးအစားပေါ်မူတည်သလားဆိုတာကို ဆုံးဖြတ်နိုင်သလို၊ saturated နဲ့ unsaturated fats တွေကြားမှာ အဆီစုပ်ယူမှု ကွာခြားသလားဆိုတာကို သိနိုင်ဖို့ a priori contrast တစ်ခုကိုလည်း ကြိုတင်သတ်မှတ်နိုင်ပါတယ်။

Statistics (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များ)

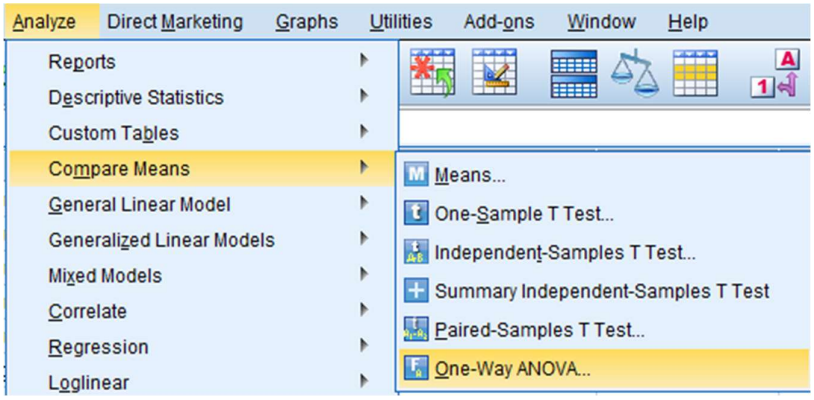
အုပ်စုတစ်ခုချင်းစီအတွက်- cases အရေအတွက်၊ mean၊ standard deviation၊ standard error of the mean၊ minimum၊ maximum၊ mean အတွက် 95% confidence interval နဲ့၊ one-way ANOVA အတွက် effect size ခန့်မှန်းချက်တွေ ပါဝင်ပါတယ်။ Levene tests for homogeneity of variance, analysis-of-variance table နဲ့ dependent variable တစ်ခုချင်းစီအတွက် means တူညီမှုကို စစ်ဆေးပေးတဲ့ robust tests တွေ၊ user ကိုယ်တိုင်သတ်မှတ်ထားတဲ့ a priori contrasts တွေ ပါဝင်ပါတယ်။ ဒါ့အပြင် Bonferroni, Sidak, Tukey, Hochberg, Gabriel, Dunnett, Duncan စတဲ့ post hoc range tests တွေနဲ့ multiple comparisons (အစုံလိုက် နှိုင်းယှဉ်ချက်များ) တွေကိုပါ ရရှိနိုင်ပါတယ်။

One-Way ANOVA အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ

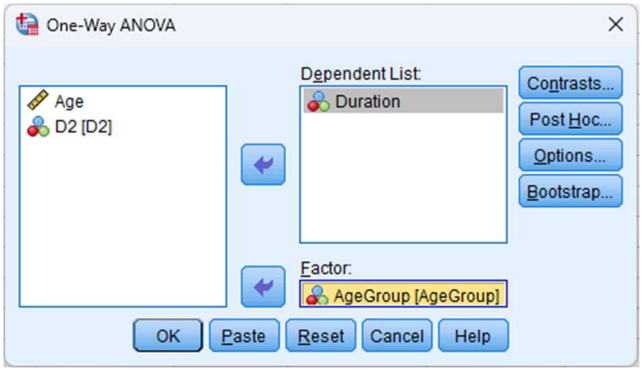
- **Data (အချက်အလက်)။** Dependent variable ဟာ အရေအတွက်ပြ (quantitative - interval level of measurement) ဖြစ်သင့်ပါတယ်။
- **Assumptions (ယူဆချက်များ)။** အုပ်စုတစ်ခုစီဟာ normal population ကနေ ထုတ်ယူထားတဲ့ independent random sample (အချင်းချင်းမမှီခိုတဲ့ ကျပန်းနမူနာ) တွေ ဖြစ်ရပါမယ်။ Data တွေက symmetric (အချိုးညီ) ဖြစ်သင့်ပေမယ့်၊ Analysis of variance ဟာ normality ကနေ သွေဖည်သွားမှုတွေကိုလည်း အထိုက်အလျောက် ခံနိုင်ရည်ရှိ (robust ဖြစ်) ပါတယ်။ အုပ်စုတွေဟာ equal variances (တူညီတဲ့ variances) ရှိတဲ့ populations တွေဆီကနေ လာတာဖြစ်သင့်ပါတယ်။ ဒီ assumption ကို စမ်းသပ်ဖို့အတွက် Levene's homogeneity-of-variance test ကို အသုံးပြုနိုင်ပါတယ်။

One-Way analysis of variance ရယူအသုံးပြုရန် အဆင့်များ

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Compare Means > One-Way ANOVA...

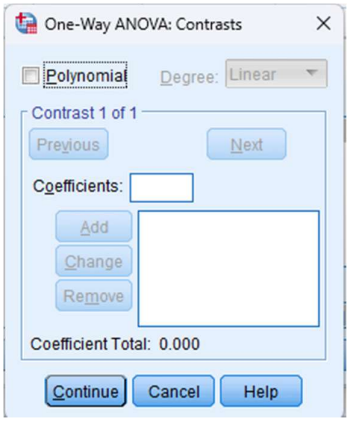


၂။ dependent variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။

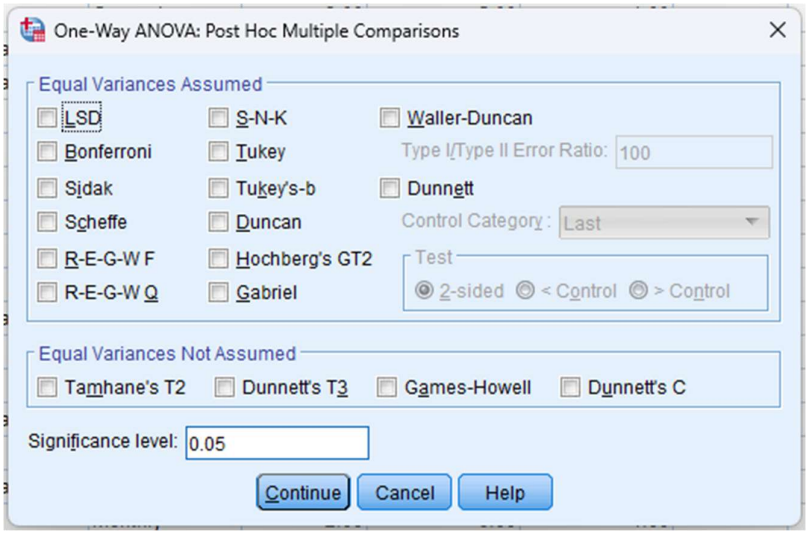


၃။ independent factor variable တစ်ခုတည်းကို ရွေးချယ်ပါ။ ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့ သင် အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

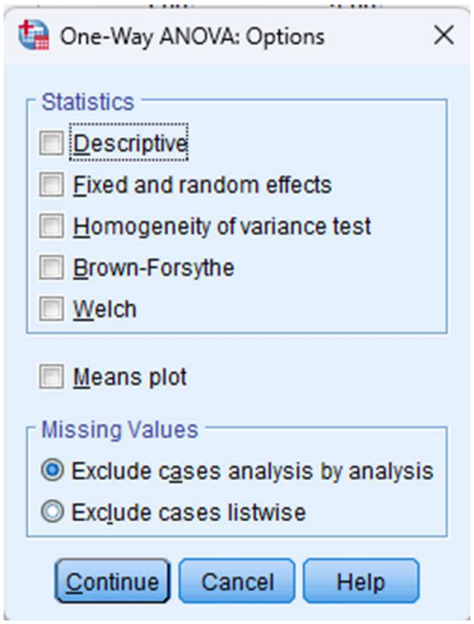
- Overall test အတွက် effect size တွက်ချက်မှုကို ထိန်းချုပ်ဖို့ **Estimate effect size for overall tests** ကို ရွေးချယ်နိုင်ပါတယ် (ရွေးချယ်လိုက်ရင် output မှာ “ANOVA Effect Sizes” ဇယား ပြသပေးမှာပါ) ။
- Between-groups sums of squares တွေကို trend components တွေအဖြစ် အပိုင်းပိုင်းခွဲဖို့ ဒါမှမဟုတ် a priori contrasts တွေ သတ်မှတ်ဖို့ **Contrasts** ကို နှိပ်နိုင်ပါတယ်။



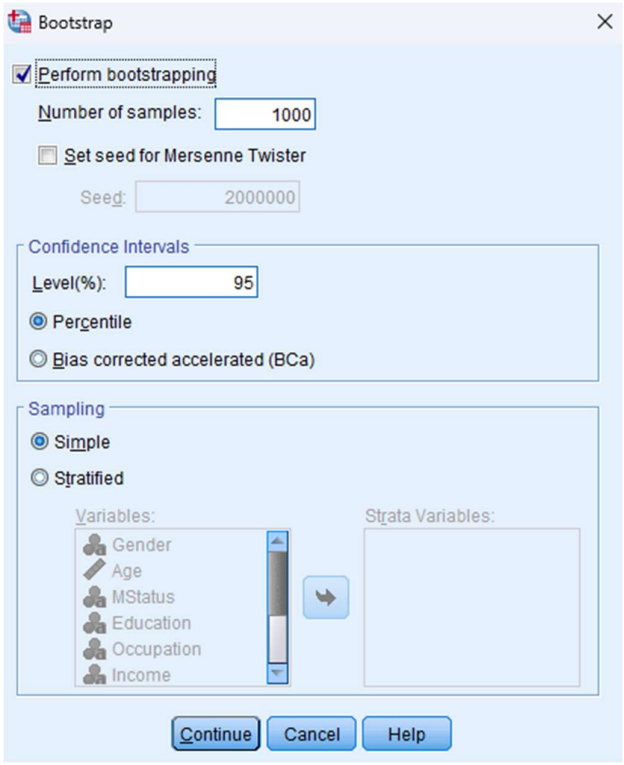
- ဘယ် means တွေ ကွာခြားနေလဲဆိုတာ ဆုံးဖြတ်ဖို့ post hoc range tests တွေနဲ့ pairwise multiple comparisons တွေကို သုံးချင်ရင် **Post Hoc** ကို နှိပ်ပါ။



- Missing data တွေကို ကိုင်တွယ်ပုံနဲ့ confidence interval level ကို ထိန်းချုပ်ဖို့ **Options** ကို နှိပ်ပါ။



- ခန့်မှန်းချက်တွေအတွက် ပိုမိုခိုင်မာတဲ့ standard errors တွေနဲ့ confidence intervals တွေ ရယူတွက်ချက်ဖို့ **Bootstrap** ကို နှိပ်နိုင်ပါတယ်။



One-Way ANOVA Contrasts (ယှဉ်တွဲကွာခြားချက်များ)

သင်က between-groups sums of squares တွေကို trend components တွေအနေနဲ့ အပိုင်းပိုင်းခွဲနိုင်သလို a priori contrasts တွေကိုလည်း သတ်မှတ်ပေးနိုင်ပါတယ်။

- **Polynomial (ပါဝါကိန်းစဉ်ဆက်)။** Factor variable ရဲ့ အစဉ်လိုက် levels တွေတစ်လျှောက် dependent variable ရဲ့ trend (ပြောင်းလဲမှုလားရာ) ကို စမ်းသပ်နိုင်ပါတယ် (ဥပမာ- ရရှိထားတဲ့ ဘွဲ့ဒီဂရီ အဆင့်တွေအလိုက် လစာငွေရဲ့ linear trend ကို စမ်းသပ်တာမျိုးပါ) ။ 1st ကနေ 5th degree polynomial အထိ ရွေးချယ်နိုင်ပါတယ်။
- **Coefficients (မြှောက်ဖော်ကိန်းများ)။** t statistic နဲ့ စမ်းသပ်မယ့် user သတ်မှတ်ထားတဲ့ a priori contrasts တွေ ဖြစ်ပါတယ်။ Factor variable ရဲ့ အုပ်စုတစ်ခုချင်းစီအတွက် coefficient တစ်ခုစီ ရိုက်ထည့်ပြီး Add ကို နှိပ်ပါ။ (မှတ်ချက် - Coefficients တွေရဲ့ အစီအစဉ်က အုပ်စုတန်ဖိုးတွေရဲ့ အငယ်မှအကြီး အစီအစဉ်နဲ့ တိုက်ရိုက်သက်ဆိုင်နေလို့ အရေးကြီးပါတယ်။ အများစုသော အခြေအနေတွေမှာ coefficients တွေအားလုံး ပေါင်းလိုက်ရင် သုည (0) ဖြစ်ရပါမယ်) ။
- **Estimate effect size for contrasts။** Contrast တွေအတွက် effect size တွက်ချက်မှုကို ထိန်းချုပ်ပေးပါတယ်။ ဒီနေရာမှာ standardizer အဖြစ် အုပ်စုအားလုံးရဲ့ pooled standard deviation ကို သုံးမလား (ဒါက default ပါ) ၊ သို့မဟုတ် contrast မှာပါဝင်တဲ့ အုပ်စုတွေရဲ့ pooled standard deviation ကိုပဲ သုံးမလား ဆိုတာ ရွေးချယ်နိုင်ပါတယ်။

One-Way ANOVA Post Hoc Tests (စမ်းသပ်ပြီးနောက် နှိုင်းယှဉ်ချက်များ)

Means တွေကြားမှာ ကွာခြားချက်တွေ ရှိနေပြီလို့ ဆုံးဖြတ်ပြီးသွားတဲ့အခါ၊ ဘယ် means တွေ ကွာခြားနေသလဲဆိုတာကို သိဖို့ သုံးပါတယ်။ Range tests တွေက အချင်းချင်း ကွာခြားမှုမရှိတဲ့ တူညီသော means အုပ်စုလေးတွေ (homogeneous subsets) ကို ခွဲခြားပေးပြီး၊ Pairwise multiple comparisons ကတော့ means တွေကို တစ်စုံချင်းစီ ယှဉ်စစ်ပေးပါတယ်။

- **Equal Variances Assumed (Variances များ တူညီသည်ဟု ယူဆသောအခါ)-** Tukey's honestly significant difference, Hochberg's GT2, Gabriel, နဲ့ Scheffé တို့က multiple comparison ရော range tests ပါ ဖြစ်ပါတယ်။ အခြားရနိုင်တဲ့ tests တွေကတော့ LSD (အစုံလိုက် t tests သုံးသည်), Bonferroni (error rate ကို ထိန်းချုပ်ပေးသည်), Sidak, R-E-G-W F, R-E-G-W Q, S-N-K, Tukey's b, Duncan, Waller-Duncan နဲ့ Dunnett (control အုပ်စုနဲ့ ယှဉ်စစ်သည့် test) တို့ ဖြစ်ကြပါတယ်။

- **Equal Variances Not Assumed (Variances များ မတူညီဟု ယူဆသောအခါ)-** Variances တွေ မတူညီတဲ့အခါမှာ သုံးသင့်တဲ့ tests တွေဖြစ်ပြီး Tamhane's T2, Dunnett's T3, Games-Howell, နဲ့ Dunnett's C တို့ ပါဝင်ပါတယ်။

Null Hypothesis test ကတော့ post hoc test အတွက် significance level (alpha) ကို ပုံမှန်အတိုင်း သုံးမလား၊ သီးသန့် ပြောင်းလဲသတ်မှတ်မလား ဆိုတာကို ရွေးချယ်ပေးတာပါ။

One-Way ANOVA Options (One-Way ANOVA ဆိုင်ရာ ရွေးချယ်စရာများ)

Statistics (စာရင်းအင်းများ)

အောက်ဖော်ပြပါ အချက်တွေထဲက တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်နိုင်ပါတယ်-

- **Descriptive (ဖော်ပြချက်ဆိုင်ရာ) -** အုပ်စုတစ်ခုချင်းစီမှာရှိတဲ့ dependent variable တစ်ခုချင်းစီအတွက် cases အရေအတွက်၊ mean (ပျမ်းမျှကိန်း)၊ standard deviation (စံသွေဖည်ကိန်း)၊ standard error of the mean (ပျမ်းမျှကိန်း၏ စံအမှား) ၊ minimum (အနည်းဆုံး)၊ maximum (အများဆုံး) နဲ့ 95% confidence intervals တွေကို တွက်ချက်ပေးပါတယ်။
- **Fixed and random effects (ပုံသေနှင့် ကျပန်း သက်ရောက်မှုများ) -** Fixed-effects model အတွက် standard deviation, standard error နဲ့ 95% confidence interval တွေကို ပြသပေးပြီး၊ random-effects model အတွက်ဆိုရင်တော့ standard error, 95% confidence interval နဲ့ အစိတ်အပိုင်းများအကြား ကွဲလွဲမှု (between-components variance) ခန့်မှန်းချက်တွေကို ပြသပေးပါတယ်။
- **Homogeneity of variance test (ကွဲလွဲကိန်းများ တူညီမှုရှိမရှိ စမ်းသပ်ခြင်း) -** အုပ်စုတွေရဲ့ variances တွေ တူညီမှု ရှိမရှိ စမ်းသပ်ဖို့ Levene statistic ကို တွက်ချက်ပေးပါတယ်။ ဒီ test ဟာ data တွေ normal distribution ဖြစ်ရမယ်ဆိုတဲ့ ယူဆချက် (assumption of normality) အပေါ်မှာ မှီခိုနေခြင်း မရှိပါဘူး။
- **Brown-Forsythe -** အုပ်စုတွေရဲ့ means တွေ တူညီမှု ရှိမရှိ စမ်းသပ်ဖို့ Brown-Forsythe statistic ကို တွက်ချက်ပေးပါတယ်။ Equal variances ဖြစ်တယ်ဆိုတဲ့ ယူဆချက် မမှန်ကန်တော့တဲ့အခါမျိုးမှာ ဒီ statistic က F statistic ထက် ပိုပြီး သုံးလို့ကောင်းပါတယ်။
- **Welch -** အုပ်စုတွေရဲ့ means တွေ တူညီမှု ရှိမရှိ စမ်းသပ်ဖို့ Welch statistic ကို တွက်ချက်ပေးပါတယ်။ ဒါလည်းပဲ Equal variances ဖြစ်တယ်ဆိုတဲ့ ယူဆချက် မမှန်ကန်တော့တဲ့အခါ F statistic ထက် ပိုပြီး သုံးလို့ကောင်းတဲ့ statistic ဖြစ်ပါတယ်။

Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)

Missing values တွေကို ဘယ်လို ကိုင်တွယ်ဖြေရှင်းမလဲဆိုတာကို ထိန်းချုပ်ပေးပါတယ်။

- **Exclude cases analysis by analysis -** ပေးထားတဲ့ analysis တစ်ခုအတွက် dependent variable မှာဖြစ်စေ၊ factor variable မှာဖြစ်စေ missing value ရှိနေတဲ့ case မျိုးကို အဲဒီ analysis မှာ အသုံးပြုမှာ မဟုတ်ပါဘူး။ ဒါ့အပြင် factor variable အတွက် သတ်မှတ်ထားတဲ့ range (အကွာအဝေး) ပြင်ပကို ရောက်နေတဲ့ case ကိုလည်း အသုံးပြုမှာ မဟုတ်ပါဘူး။
- **Exclude cases listwise -** Factor variable မှာဖြစ်စေ၊ main dialog box ရဲ့ dependent list ထဲမှာပါတဲ့ dependent variable တစ်ခုခုမှာဖြစ်စေ missing values ရှိနေတဲ့ cases တွေကို analysis အားလုံးကနေ ဖယ်ထုတ်လိုက်မှာ ဖြစ်ပါတယ်။ အကယ်၍ သင်က dependent variables အများကြီးကို သတ်မှတ်ထားတာ မဟုတ်ဘူးဆိုရင်တော့ ဒီ option က ဘာသက်ရောက်မှုမှ ရှိမှာ မဟုတ်ပါဘူး။

Confidence Interval (ယုံကြည်မှု အတိုင်းအတာ)

ပုံမှန်အားဖြင့် mean နဲ့ hypothesized test value တို့ကြားက ကွာခြားချက်အတွက် 95% confidence interval ကို ပြသပေးပါတယ်။ တခြား confidence level ကို လိုချင်တယ်ဆိုရင်တော့ 1 ကနေ 99 ကြား တန်ဖိုးတစ်ခုခုကို ရိုက်ထည့်နိုင်ပါတယ်။

Means plot (ပျမ်းမျှကိန်းများ ဇယား)

Subgroup means (factor variable ရဲ့ တန်ဖိုးတွေက သတ်မှတ်ပေးလိုက်တဲ့ အုပ်စုတစ်ခုစီရဲ့ means) တွေကို ပုံဖော်ပြသပေးတဲ့ chart တစ်ခုကို ပြသပေးပါတယ်။

ONEWAY Command Additional Features (ONEWAY Command ၏ အခြားလုပ်ဆောင်ချက်များ)

Command syntax language ကိုအသုံးပြုပြီး အောက်ပါတို့ကိုလည်း လုပ်ဆောင်နိုင်ပါသေးတယ်-

- Fixed- နှင့် random-effects statistics တွေကို ရယူနိုင်ပါတယ်။ Fixed-effects model အတွက် standard deviation, standard error of the mean နဲ့ 95% confidence intervals တွေကို ရယူနိုင်သလို၊ random-effects model အတွက် standard error, 95% confidence intervals နဲ့ between-components variance ခန့်မှန်းချက်တွေကို ရယူနိုင်ပါတယ် (STATISTICS=EFFECTS ကို သုံးပြီး) ။

- Least significance difference, Bonferroni, Duncan နဲ့ Scheffé multiple comparison tests တွေအတွက် alpha levels (အယ်လ်ဖာ အဆင့်များ) ကို သတ်မှတ်ပေးနိုင်ပါတယ် (RANGES subcommand ကို သုံးပြီး) ။
- Means, standard deviations နဲ့ frequencies တွေပါဝင်တဲ့ matrix တစ်ခုကို ရေးသားနိုင်သလို၊ means, frequencies, pooled variances နဲ့ pooled variances အတွက် degrees of freedom တွေပါဝင်တဲ့ matrix ကိုလည်း ဖတ်ယူနိုင်ပါတယ်။ ဒီ matrices တွေကို one-way analysis of variance ရယူဖို့အတွက် raw data တွေနေရာမှာ အစားထိုး အသုံးပြုနိုင်ပါတယ် (MATRIX subcommand ကို သုံးပြီး) ။

(ပြီးပြည့်စုံတဲ့ syntax အချက်အလက်တွေအတွက် Command Syntax Reference ကို ကြည့်ရှုနိုင်ပါတယ်။)

Bivariate Correlations (ပြောင်းလဲကိန်းနှစ်ခုကြား ဆက်စပ်မှု)

Bivariate Correlations procedure ဟာ Pearson's correlation coefficient, Spearman's rho နဲ့ Kendall's tau-b တို့ကို သူတို့ရဲ့ significance levels (အရေးပါမှုအဆင့်များ) နဲ့အတူ တွက်ချက်ပေးပါတယ်။ Correlations ဆိုတာ variables တွေ ဒါမှမဟုတ် rank အစဉ်လိုက်တွေ ဘယ်လို ဆက်စပ်နေသလဲဆိုတာကို တိုင်းတာတာပါ။ Correlation coefficient ကို မတွက်ချက်ခင်မှာ သင့် data တွေထဲမှာ (မှားယွင်းတဲ့ ရလဒ်တွေ ထွက်စေနိုင်တဲ့) outliers (လွန်ကဲတန်ဖိုးများ) ပါဝင်နေသလား နဲ့ linear ဆက်စပ်မှု (မျဉ်းဖြောင့်ဆက်စပ်မှု) ရှိမရှိကို အရင်ဆုံး စစ်ဆေးသင့်ပါတယ်။ Pearson's correlation coefficient ဟာ မျဉ်းဖြောင့်ဆက်စပ်မှု (linear association) ကို တိုင်းတာတဲ့ အတိုင်းအတာတစ်ခုပါ။ Variable နှစ်ခုဟာ ပြီးပြည့်စုံစွာ ဆက်စပ်နေနိုင်ပေမယ့်၊ သူတို့ရဲ့ ဆက်စပ်မှုက linear မဟုတ်ဘူးဆိုရင်တော့ သူတို့ရဲ့ ဆက်စပ်မှုကို တိုင်းတာဖို့ Pearson's correlation coefficient က သင့်လျော်တဲ့ statistic မဟုတ်ပါဘူး။

Pearson နဲ့ Spearman တို့အတွက် Confidence interval ဆိုင်ရာ setting တွေကိုလည်း အသုံးပြုနိုင်ပါတယ်။

Example (ဥပမာ)

ဘတ်စကက်ဘောအသင်းတစ်သင်း နိုင်တဲ့ ပွဲအရေအတွက်နဲ့ တစ်ပွဲမှာ ပျမ်းမျှရရှိတဲ့ အမှတ်အရေအတွက်က ဆက်စပ်မှုရှိသလား။ Scatterplot (အစက်ချဇယား) ဆွဲကြည့်တဲ့အခါ မျဉ်းဖြောင့် ဆက်စပ်မှု (linear relationship) ရှိနေတာကို တွေ့ရပါတယ်။ ၁၉၉၄-၁၉၉၅ NBA ရာသီက data တွေကို ခွဲခြမ်းစိတ်ဖြာကြည့်တဲ့အခါ Pearson's correlation coefficient (0.581) ရရှိပြီး 0.01 အဆင့်မှာ significant ဖြစ်တာ တွေ့ရပါတယ်။ တစ်ရာသီအတွင်း နိုင်ပွဲများလေ၊ ပြိုင်ဘက်က ရတဲ့ အမှတ်နည်းလေပဲလို့ သင် ယူဆကောင်း ယူဆပါလိမ့်မယ်။ ဒီ variables နှစ်ခုဟာ ပြောင်းပြန်ဆက်စပ်မှု (negatively correlated, -0.401) ရှိပြီး 0.05 အဆင့်မှာ significant ဖြစ်တယ်ဆိုတာကို တွေ့ရှိရမှာပါ။

Statistics (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များ)

Variable တစ်ခုစီအတွက်- missing မဖြစ်တဲ့ cases အရေအတွက်၊ mean နဲ့ standard deviation တို့ ပါဝင်ပါတယ်။ Variable စုံတွဲ တစ်စုံချင်းစီအတွက်- Pearson's correlation coefficient, Spearman's rho, Kendall's tau-b, cross-product of deviations နဲ့ covariance တို့ ပါဝင်ပါတယ်။

Bivariate Correlations အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ

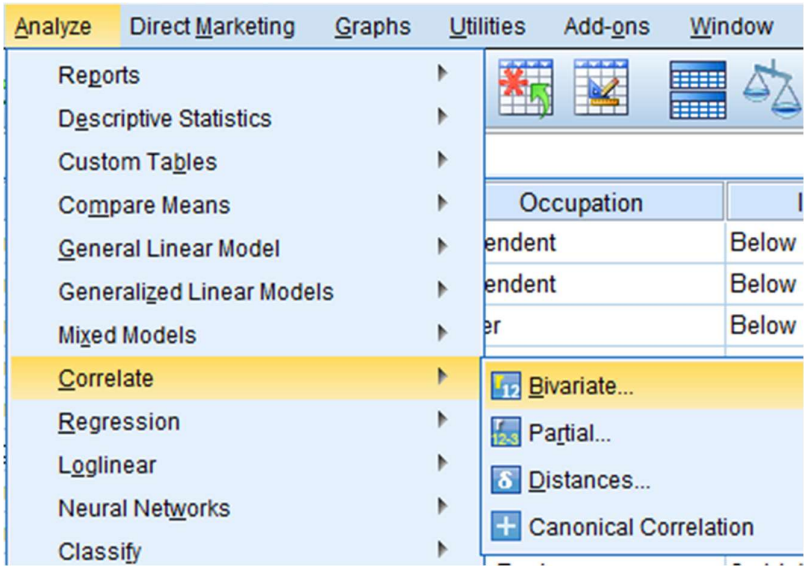
- **Data (အချက်အလက်)။** Pearson's correlation coefficient အတွက် အချိုးညီတဲ့ (symmetric) အရေအတွက်ပြ (quantitative) variables တွေကို အသုံးပြုပါ။ Spearman's

rho နဲ့ Kendall's tau-b အတွက်ကတော့ quantitative variables တွေ ဒါမှမဟုတ် အစဉ်လိုက် categories (ordered categories) တွေပါတဲ့ variables တွေကို အသုံးပြုပါ။

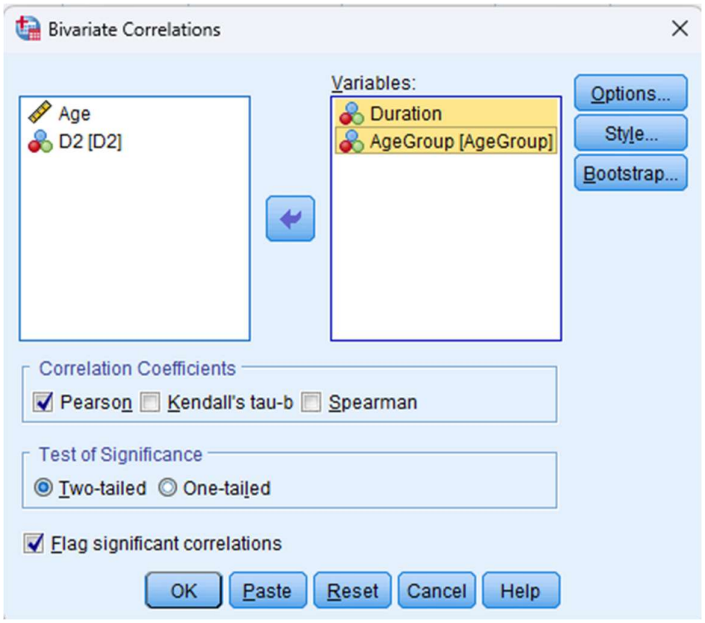
- **Assumptions (ယူဆချက်များ)။** Pearson's correlation coefficient တွက်ချက်ဖို့ variable စုံတွဲတစ်ခုစီဟာ bivariate normal (နှစ်ခုစလုံး normal distribution ဖြစ်ခြင်း) ဖြစ်တယ်လို့ ယူဆထားပါတယ်။

Bivariate Correlations ရယူအသုံးပြုရန် အဆင့်များ

menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Correlate > Bivariate...



၁။ ဂဏန်း (numeric) variable နှစ်ခု သို့မဟုတ် နှစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။



အောက်ဖော်ပြပါ ရွေးချယ်စရာ (options) တွေကိုလည်း အသုံးပြုနိုင်ပါတယ်-

- **Correlation Coefficients (ဆက်စပ်မှု ကိန်းဂဏန်းများ)။** Normally distributed ဖြစ်တဲ့ (ပုံမှန်ဖြန့်ကျက်မှုရှိတဲ့) အရေအတွက်ပြ variables တွေအတွက် Pearson correlation coefficient ကို ရွေးချယ်ပါ။ သင့် data က normal distribution မဟုတ်ဘူး ဒါမှမဟုတ် အစဉ်လိုက် categories တွေဖြစ်နေတယ်ဆိုရင်၊ rank အစဉ်လိုက်တွေကြားက ဆက်စပ်မှုကို တိုင်းတာပေးတဲ့ Kendall's tau-b ဒါမှမဟုတ် Spearman ကို ရွေးချယ်ပါ။ Correlation coefficients တန်ဖိုးတွေဟာ -1 (ပြီးပြည့်စုံတဲ့ ပြောင်းပြန်ဆက်စပ်မှု) နဲ့ +1 (ပြီးပြည့်စုံတဲ့ အတည့်ဆက်စပ်မှု) ကြားမှာ ရှိပါတယ်။ တန်ဖိုး 0 ဆိုရင်တော့ မျဉ်းဖြောင့်ဆက်စပ်မှု (linear relationship) မရှိဘူးလို့ ဆိုလိုပါတယ်။ ရလဒ်တွေကို ကောက်ချက်ချတဲ့အခါ၊ correlation က significant ဖြစ်တယ်ဆိုရုံနဲ့ အကြောင်းကျိုးဆက်စပ်မှု (cause-and-effect conclusions) ရှိတယ်လို့ ကောက်ချက်မချမိဖို့ သတိထားရပါမယ်။

- **Test of Significance (အရေးပါမှု စမ်းသပ်ခြင်း) ။** Two-tailed ဒါမှမဟုတ် One-tailed probabilities ကို ရွေးချယ်နိုင်ပါတယ်။ ဆက်စပ်မှုရဲ့ ဦးတည်ချက်ကို ကြိုတင်သိထားတယ်ဆိုရင် One-tailed ကို ရွေးချယ်ပါ။ မသိဘူးဆိုရင်တော့ Two-tailed ကို ရွေးချယ်ပါ။

- **Flag significant correlations (အရေးပါသော ဆက်စပ်မှုများကို မှတ်သားခြင်း) ။** 0.05 level မှာ significant ဖြစ်တဲ့ correlation coefficients တွေကို ခရေပွင့် တစ်ပွင့် (*) နဲ့ ပြသပြီး၊ 0.01 level မှာ significant ဖြစ်တာတွေကိုတော့ ခရေပွင့် နှစ်ပွင့် (**) နဲ့ ပြသပေးမှာ ဖြစ်ပါတယ်။

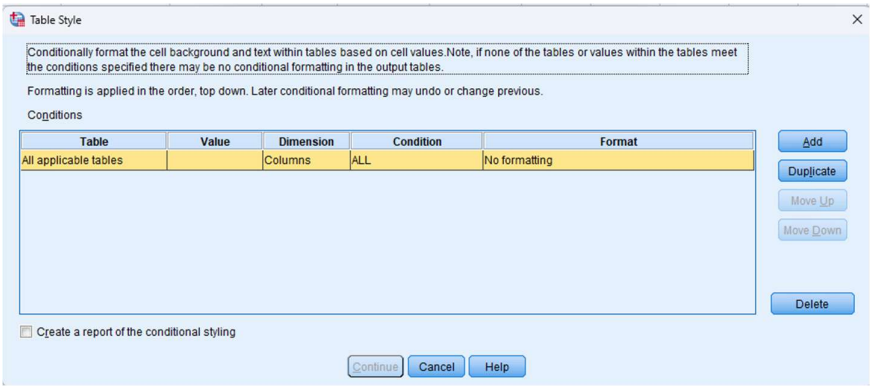
- **Show only the lower triangle (အောက်ဘက် တြိဂံကိုသာ ပြသရန်) ။** ဒါကို ရွေးချယ်လိုက်ရင် correlation matrix ဇယားရဲ့ အောက်ဘက်တြိဂံပုံ အပိုင်းကိုပဲ output မှာ ပြသပေးမှာပါ။ မရွေးချယ်ရင်တော့ ဇယားအပြည့်ကို ပြသပေးမှာပါ။ ဒီ setting က ဇယားကို APA style လမ်းညွှန်ချက်တွေနဲ့ ကိုက်ညီစေဖို့ ကူညီပေးပါတယ်။

- **Show diagonal (ဒေါင့်ဖြတ်မျဉ်း ပြသရန်) ။** ဒါကို ရွေးချယ်ရင်တော့ အောက်ဘက်တြိဂံပုံ အပိုင်းနဲ့အတူ ဒေါင့်ဖြတ်မျဉ်း (diagonal) တန်ဖိုးတွေကိုပါ ပြသပေးမှာ ဖြစ်ပါတယ် (ဒါဟာလည်း APA style နဲ့ ကိုက်ညီစေဖို့ပါ) ။

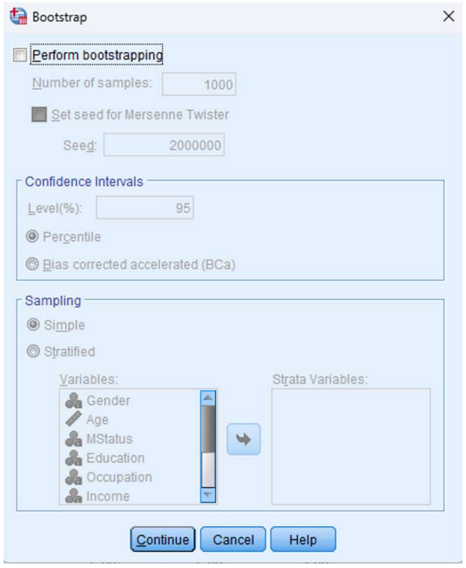
၂။ ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များအနေနဲ့ သင် အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- Pearson correlation statistics တွေနဲ့ missing values ကိုင်တွယ်ပုံကို သတ်မှတ်ဖို့ **Options...** ကို နှိပ်နိုင်ပါတယ်။

- သီးခြား အခြေအနေတွေအပေါ် မူတည်ပြီး pivot tables တွေရဲ့ properties တွေကို အလိုအလျောက် ပြောင်းလဲပေးဖို့ **Style...** ကို နှိပ်နိုင်ပါတယ်။



- Mean, median, proportion, odds ratio, correlation coefficient ဒါမှမဟုတ် regression coefficient လိုမျိုး ခန့်မှန်းချက်တွေအတွက် ပိုမိုခိုင်မာတဲ့ (robust) standard errors တွေနဲ့ confidence intervals တွေကို ရယူတွက်ချက်ဖို့ **Bootstrap...** ကို နှိပ်နိုင်ပါတယ်။



- Confidence intervals ခန့်မှန်းတွက်ချက်မှုအတွက် ရွေးချယ်စရာတွေကို သတ်မှတ်ဖို့ **Confidence Interval...** ကို နှိပ်နိုင်ပါတယ်။

Bivariate Correlations Options (Bivariate Correlations ဆိုင်ရာ ရွေးချယ်စရာများ)

Statistics (စာရင်းအင်းများ)

Pearson correlations အတွက် အောက်ပါအချက်တွေထဲက တစ်ခု ဒါမှမဟုတ် နှစ်ခုလုံးကို ရွေးချယ်နိုင်ပါတယ်-

- **Means and standard deviations (ပျမ်းမျှကိန်းများနှင့် စံသွေဖည်ကိန်းများ)** ။ Variable တစ်ခုချင်းစီအတွက် ပြသပေးပါတယ်။ Missing မဖြစ်တဲ့ (မပျောက်ဆုံးတဲ့) တန်ဖိုးတွေရှိတဲ့ cases အရေအတွက်ကိုလည်း ပြသပေးပါတယ်။ သင်သတ်မှတ်ထားတဲ့ missing values အပြင်အဆင် (setting) ဘယ်လိုပဲဖြစ်နေပါစေ၊ missing values တွေကိုတော့ variable တစ်ခုချင်းစီအလိုက် သီးခြားစီ ကိုင်တွယ်ဖြေရှင်းပေးသွားမှာပါ။

- **Cross-product deviations and covariances**။ Variable စုံတွဲ တစ်စုံချင်းစီအတွက် ပြသပေးပါတယ်။ Cross-product of deviations ဆိုတာ mean-corrected variables တွေရဲ့ မြှောက်လဒ်တွေ ပေါင်းထားတဲ့ ပေါင်းလဒ်နဲ့ ညီမျှပါတယ်။ ဒါဟာ Pearson correlation coefficient ရဲ့ ပိုင်းဝေ (numerator) ဖြစ်ပါတယ်။ Covariance ဆိုတာကတော့ variable နှစ်ခုကြားက ဆက်စပ်မှုကို စံသတ်မှတ်မထားဘဲ (unstandardized) တိုင်းတာတဲ့ အတိုင်းအတာတစ်ခုဖြစ်ပြီး၊ cross-product deviation ကို cases အရေအတွက်အောက် တစ်လျော့တဲ့ကိန်း (N-1) နဲ့ စားထားတဲ့ ရလဒ်နဲ့ ညီမျှပါတယ်။

Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)

အောက်ပါတို့ထဲက တစ်ခုကို ရွေးချယ်နိုင်ပါတယ်-

- **Exclude cases pairwise**။ Correlation coefficient တွက်ချက်ဖို့ variable စုံတွဲတစ်ခုမှာ တစ်ခုဖြစ်စေ၊ နှစ်ခုလုံးမှာဖြစ်စေ missing values ရှိနေတဲ့ cases တွေကို analysis ကနေ ဖယ်ထုတ်လိုက်ပါတယ်။ Coefficient တစ်ခုချင်းစီကို အဲဒီ သီးခြား variable စုံတွဲအတွက် မှန်ကန်တဲ့ (valid) ကုဒ်တွေရှိတဲ့ cases အားလုံးပေါ် အခြေခံပြီး တွက်ချက်တဲ့အတွက်၊ တွက်ချက်မှုတိုင်းမှာ ရနိုင်သမျှ အများဆုံး သတင်းအချက်အလက်ကို အသုံးပြုသွားမှာ ဖြစ်ပါတယ်။ ဒါပေမယ့် ဒီလိုလုပ်ခြင်းအားဖြင့် ထွက်ပေါ်လာတဲ့ coefficients တွေဟာ မတူညီတဲ့ cases အရေအတွက်တွေပေါ် အခြေခံပြီး တွက်ချက်ထားတာမျိုး ဖြစ်သွားနိုင်ပါတယ်။

- **Exclude cases listwise**။ Variable တစ်ခုခုမှာ missing values ရှိနေတဲ့ cases မှန်သမျှကို correlations အားလုံးကနေ ဖယ်ထုတ်လိုက်ပါတယ်။

Bivariate Correlations Confidence Interval (Bivariate Correlations ယုံကြည်မှု အတိုင်းအတာ)

Confidence Interval dialog က confidence intervals တွေကို ခန့်မှန်းတွက်ချက်ဖို့ ရွေးချယ်စရာတွေ ထောက်ပံ့ပေးပါတယ်။ ဒီ dialog ကို Bivariate Correlations dialog ထဲမှာ Pearson, Kendall's tau-b, ဒါမှမဟုတ် Spearman ကို ရွေးချယ်ထားတဲ့အခါမှသာ အသုံးပြုနိုင်မှာပါ။

- **Estimate confidence interval of bivariate correlation parameter**။ Bivariate correlation parameter ရဲ့ confidence interval ခန့်မှန်းတွက်ချက်မှုကို ထိန်းချုပ်ပေးပါတယ်။ သူ့ကို ရွေးချယ်ထားရင် confidence interval ခန့်မှန်းတွက်ချက်မှုကို လုပ်ဆောင်ပေးမှာ ဖြစ်ပါတယ်။
- **Confidence interval (%)** ။ တွက်ထုတ်ပေးမယ့် confidence intervals အားလုံးအတွက် confidence level ကို သတ်မှတ်ပေးပါတယ်။ 0 နဲ့ 100 ကြားက ဂဏန်းတန်ဖိုးတစ်ခုကို သတ်မှတ်ပေးပါ။ ပုံမှန်အားဖြင့် (Default) 95 ကို သတ်မှတ်ထားပါတယ်။
- **Pearson Correlation**။ *Apply the bias adjustment* ဆိုတဲ့ အပြင်အဆင်လေးက bias ချိန်ညှိမှု (bias adjustment) ကို အသုံးပြုမပြု ထိန်းချုပ်ပေးပါတယ်။ ပုံမှန်အားဖြင့် ဒီအပြင်အဆင်ကို ရွေးချယ်မထားပါဘူး၊ ဒါကြောင့် bias term ကို ထည့်သွင်းစဉ်းစားမှာ မဟုတ်ပါဘူး။ တကယ်လို့ ရွေးချယ်ထားမယ်ဆိုရင်တော့ confidence limits တွေ ခန့်မှန်းတွက်ချက်ရာမှာ bias adjustment ကို ထည့်သွင်းအသုံးပြုပေးမှာ ဖြစ်ပါတယ်။ ဒီအပြင်အဆင်ကို Bivariate Correlations dialog ထဲမှာ Pearson ကို ရွေးချယ်ထားမှသာ အသုံးပြုနိုင်မှာပါ။
- **Spearman Correlation**။ ဒီအပြင်အဆင်ကို Bivariate Correlations dialog ထဲမှာ Spearman ကို ရွေးချယ်ထားတဲ့အခါမှသာ အသုံးပြုနိုင်ပြီး၊ Spearman Correlation variance ကို အောက်ပါနည်းလမ်းတွေကတစ်ဆင့် ခန့်မှန်းတွက်ချက်ဖို့ ရွေးချယ်စရာတွေ ထောက်ပံ့ပေးပါတယ်-
    - Fieller, Hartley and Pearson
    - Bonett and Wright
    - Coruso and Cliff

Partial Correlations (တစ်စိတ်တစ်ပိုင်း ဆက်စပ်မှုများ)

Partial Correlations procedure ဟာ variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုတဲ့ အခြား variables တွေရဲ့ သက်ရောက်မှု (effects) တွေကို ထိန်းချုပ် (control) ထားရင်းနဲ့၊ variables နှစ်ခုကြားက မျဉ်းဖြောင့်ဆက်စပ်မှု (linear relationship) ကို ဖော်ပြပေးတဲ့ partial correlation coefficients တွေကို တွက်ချက်ပေးပါတယ်။ Correlations ဆိုတာ မျဉ်းဖြောင့်ဆက်စပ်မှုကို တိုင်းတာတာပါ။ Variables နှစ်ခုက ပြီးပြည့်စုံစွာ ဆက်စပ်နေနိုင်ပေမယ့်၊ သူတို့ရဲ့ ဆက်စပ်မှုက မျဉ်းဖြောင့် (linear) မဟုတ်ဘူးဆိုရင်တော့ correlation coefficient က သူတို့ရဲ့ဆက်စပ်မှုကို တိုင်းတာဖို့ သင့်လျော်တဲ့ statistic မဟုတ်ပါဘူး။

Example (ဥပမာ)

ကျန်းမာရေးစောင့်ရှောက်မှု ရန်ပုံငွေ (healthcare funding) နဲ့ ရောဂါဖြစ်ပွားမှုနှုန်း (disease rates) ကြားမှာ ဆက်စပ်မှု ရှိပါသလား။ ဒီလိုဆက်စပ်မှုမျိုးဟာ ပြောင်းပြန်ဆက်စပ်မှု (negative correlation - ရန်ပုံငွေများရင် ရောဂါနည်းရမယ်) ဖြစ်မယ်လို့ သင် မျှော်လင့်ထားနိုင်ပေမယ့်၊ လေ့လာမှုတစ်ခုရဲ့ အစီရင်ခံစာအရ သိသာထင်ရှားတဲ့ အတည့်ဆက်စပ်မှု (positive correlation) ကို တွေ့ရပါတယ် (ဆိုလိုတာက ကျန်းမာရေးရန်ပုံငွေ တိုးလာတာနဲ့အမျှ ရောဂါဖြစ်ပွားမှုနှုန်းကလည်း မြင့်တက်လာပုံရနေပါတယ်)။ ဒါပေမယ့် ကျန်းမာရေးစောင့်ရှောက်ပေးသူတွေဆီ လာရောက်ပြသတဲ့နှုန်း (rate of visits to healthcare providers) ကိုသာ control လုပ် (ထိန်းချုပ်) ကြည့်လိုက်မယ်ဆိုရင်တော့၊ အစက တွေ့ခဲ့ရတဲ့ positive correlation ကြီးက လုံးဝနီးပါး ပျောက်ကွယ်သွားတာကို တွေ့ရပါလိမ့်မယ်။ ကျန်းမာရေးရန်ပုံငွေနဲ့ ရောဂါဖြစ်ပွားမှုနှုန်းတွေဟာ အတည့်ဆက်စပ်နေတယ်လို့ ထင်ရခြင်းအကြောင်းရင်းက၊ ရန်ပုံငွေတွေ တိုးလာတဲ့အခါ လူတွေက ကျန်းမာရေးစောင့်ရှောက်မှုကို ပိုမိုလက်လှမ်းမီလာကြပြီး ဆရာဝန်တွေနဲ့ ဆေးရုံတွေကနေ ရောဂါတွေကို ပိုပြီး အစီရင်ခံလာနိုင်လို့ပဲ ဖြစ်ပါတယ်။

Statistics (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များ)

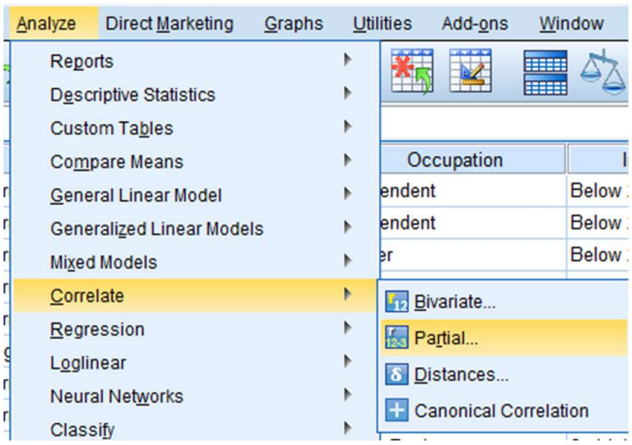
Variable တစ်ခုစီအတွက်- missing မဖြစ်တဲ့ cases အရေအတွက်၊ mean နဲ့ standard deviation တို့ ပါဝင်ပါတယ်။ Degrees of freedom နဲ့ significance levels (အရေးပါမှုအဆင့်များ) နဲ့တကွ Partial နဲ့ zero-order correlation matrices တွေကိုလည်း ထုတ်ပေးပါတယ်။

Partial Correlations အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ

- **Data (အချက်အလက်)။** အချိုးညီတဲ့ (symmetric) အရေအတွက်ပြ (quantitative) variables တွေကို အသုံးပြုပါ။
- **Assumptions (ယူဆချက်များ)။** Partial Correlations procedure က variable စုံတွဲတစ်ခုစီဟာ bivariate normal (နှစ်ခုစလုံး normal distribution ဖြစ်ခြင်း) ဖြစ်တယ်လို့ ယူဆထားပါတယ်။

Obtaining Partial Correlations (Partial Correlations ရယူအသုံးပြုရန် အဆင့်များ)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Correlate > Partial...



၂။ Partial correlations တွက်ချက်ချင်တဲ့ ဂဏန်း (numeric) variable နှစ်ခု ဒါမှမဟုတ် နှစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။

၃။ ထိန်းချုပ်မယ့် (control) numeric variables တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။

အောက်ဖော်ပြပါ ရွေးချယ်စရာ (options) တွေကိုလည်း အသုံးပြုနိုင်ပါတယ်-

- **Test of Significance (အရေးပါမှု စမ်းသပ်ခြင်း)** ။ Two-tailed ဒါမှမဟုတ် One-tailed probabilities ကို ရွေးချယ်နိုင်ပါတယ်။ ဆက်စပ်မှုရဲ့ ဦးတည်ချက်ကို ကြိုတင်သိထားတယ်ဆိုရင် One-tailed ကို ရွေးချယ်ပါ။ မသိဘူးဆိုရင်တော့ Two-tailed ကို ရွေးချယ်ပါ။
- **Display actual significance level (အမှန်တကယ် အရေးပါမှုအဆင့်ကို ပြသခြင်း)**။ ပုံမှန်အားဖြင့် correlation coefficient တစ်ခုချင်းစီအတွက် probability နဲ့ degrees of freedom တွေကို ပြသပေးပါတယ်။ တကယ်လို့ သင်က ဒီ option ကို အမှန်ခြစ်ဖြုတ်လိုက်မယ်ဆိုရင်၊ 0.05 level မှာ significant ဖြစ်တဲ့ coefficients တွေကို ခရေပွင့် တစ်ပွင့် (*) နဲ့ ပြပြီး၊ 0.01 level မှာ significant ဖြစ်တာတွေကို ခရေပွင့် နှစ်ပွင့် (**) နဲ့ ပြသပေးမှာဖြစ်ကာ degrees of freedom တွေကိုတော့ ဖျောက်ထားပါလိမ့်မယ်။ ဒီအပြင်အဆင်ဟာ partial ရော zero-order correlation matrices နှစ်ခုလုံးအပေါ် သက်ရောက်မှုရှိပါတယ်။

Partial Correlations Options (ရွေးချယ်စရာများ)

- **Statistics (စာရင်းအင်းများ)**

အောက်ပါတို့ထဲက တစ်ခု ဒါမှမဟုတ် နှစ်ခုလုံးကို ရွေးချယ်နိုင်ပါတယ်-

- ***Means and standard deviations (ပျမ်းမျှကိန်းများနှင့် စံသွေဖည်ကိန်းများ)*** - Variable တစ်ခုချင်းစီအတွက် ပြသပေးပါတယ်။ Missing မဖြစ်တဲ့ values တွေပါတဲ့ cases အရေအတွက်ကိုလည်း ပြသပေးပါတယ်။
- ***Zero-order correlations*** - Control variables တွေအပါအဝင် variables အားလုံးကြားက ရိုးရှင်းတဲ့ (simple) correlations တွေရဲ့ matrix တစ်ခုကို ပြသပေးပါတယ်။

- **Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)**

အောက်ပါတို့ထဲက တစ်ခုကို ရွေးချယ်နိုင်ပါတယ်-

- ***Exclude cases listwise*** - Control variable အပါအဝင် variable တစ်ခုခုမှာ missing values ရှိနေတဲ့ cases တွေကို တွက်ချက်မှုအားလုံးကနေ ဖယ်ထုတ်လိုက်ပါတယ်။
- ***Exclude cases pairwise*** - Partial correlations တွေ အခြေခံတွက်ချက်မယ့် zero-order correlations တွေကို တွက်ချက်ရာမှာ၊ variable စုံတွဲတစ်ခုမှာ တစ်ခုဖြစ်စေ၊ နှစ်ခုလုံးမှာဖြစ်စေ missing values ရှိနေတဲ့ case ကို အသုံးမပြုပါဘူး။ Pairwise ဖယ်ထုတ်ခြင်းက ရနိုင်သမျှ data အများဆုံးကို အသုံးပြုပါတယ်။ ဒါပေမယ့် coefficients တွေကြားမှာ အသုံးပြုသွားတဲ့ cases အရေအတွက် ကွာခြားသွားနိုင်ပါတယ်။ Pairwise deletion ကိုသုံးတဲ့အခါ၊ သီးခြား partial coefficient တစ်ခုအတွက် degrees of freedom ကို zero-order correlations တွက်ချက်ရာမှာ သုံးခဲ့တဲ့ အနည်းဆုံး cases အရေအတွက်ပေါ် အခြေခံပြီး တွက်ချက်ပါတယ်။

PARTIAL CORR Command Additional Features (အခြား လုပ်ဆောင်ချက်များ)

Command syntax language ကိုအသုံးပြုပြီး အောက်ပါတို့ကိုလည်း ထပ်မံလုပ်ဆောင်နိုင်ပါသေးတယ်-

- Zero-order correlation matrix တစ်ခုကို ဖတ်တာ ဒါမှမဟုတ် partial correlation matrix တစ်ခုကို ရေးသားတာ (MATRIX subcommand ကို သုံးပြီး) ။
- Variables စာရင်း နှစ်ခုကြားက partial correlations တွေကို ရယူတာ (VARIABLES subcommand ပေါ်က WITH keyword ကို သုံးပြီး) ။
- Multiple analyses တွေကို ရယူတာ (VARIABLES subcommands အများကြီးသုံးပြီး) ။

- Control variables နှစ်ခုရှိတဲ့အခါ တောင်းဆိုမယ့် order values တွေ (ဥပမာ - first- နဲ့ second-order partial correlations နှစ်မျိုးလုံး) ကို သတ်မှတ်တာ။
- ထပ်နေတဲ့ (redundant) coefficients တွေကို ဖျောက်ထားတာ (FORMAT subcommand သုံးပြီး)။
- Coefficients တချို့ တွက်ချက်လို့မရတဲ့အခါ simple correlations တွေရဲ့ matrix တစ်ခုကို ပြသတာ (STATISTICS subcommand သုံးပြီး) စတာတွေပဲ ဖြစ်ပါတယ်။

Linear models (လီနီယာ မော်ဒယ်များ)

Linear models တွေက target (ခန့်မှန်းချင်တဲ့ ပြောင်းလဲကိန်း) နဲ့ predictor (ခန့်မှန်းရာမှာသုံးမယ့် အချက်အလက်များ) တွေကြားက မျဉ်းဖြောင့်ဆက်စပ်မှု (linear relationships) တွေအပေါ် အခြေခံပြီး ဆက်တိုက်ပြောင်းလဲနေတဲ့ continuous target တွေကို ခန့်မှန်းတွက်ချက်ပေးပါတယ်။

Linear models တွေက ရိုးရှင်းပြီး၊ ရလဒ်တွေကို ရှင်းပြရလွယ်ကူတဲ့ သချိ်ာဖော်မြူလာတွေကို ပေးစွမ်းပါတယ်။ သူတို့ရဲ့ ဂုဏ်သတ္တိတွေကို နားလည်ရလွယ်ကူပြီး တူညီတဲ့ dataset တစ်ခုတည်းမှာပဲ Neural networks သို့မဟုတ် decision trees တွေလို တခြား model အမျိုးအစားတွေနဲ့ ယှဉ်လိုက်ရင် အများအားဖြင့် ပိုပြီး မြန်မြန်ဆန်ဆန် တည်ဆောက်နိုင်ပါတယ်။

Example (ဥပမာ)

အိမ်ပိုင်ရှင်တွေရဲ့ အာမခံတောင်းခံမှုတွေကို စစ်ဆေးဖို့ အရင်းအမြစ် (resources) ကန့်သတ်ချက်တွေ ရှိနေတဲ့ အာမခံကုမ္ပဏီတစ်ခုဟာ တောင်းခံလာမယ့် ကုန်ကျစရိတ်တွေကို ခန့်မှန်းဖို့ model တစ်ခု တည်ဆောက်ချင်တယ် ဆိုပါစို့။ ဒီ model ကို customer service စင်တာတွေမှာ အသုံးပြုခြင်းအားဖြင့်၊ ကိုယ်စားလှယ်တွေက customer နဲ့ ဖုန်းပြောနေစဉ်မှာပဲ လိုအပ်တဲ့ အချက်အလက်တွေကို ထည့်သွင်းပြီး အရင် data တွေအပေါ်မူတည်ကာ အဲဒီတောင်းခံမှုအတွက် "မျှော်မှန်းထားတဲ့" (expected) ကုန်ကျစရိတ်ကို ချက်ချင်း သိရှိနိုင်မှာ ဖြစ်ပါတယ်။

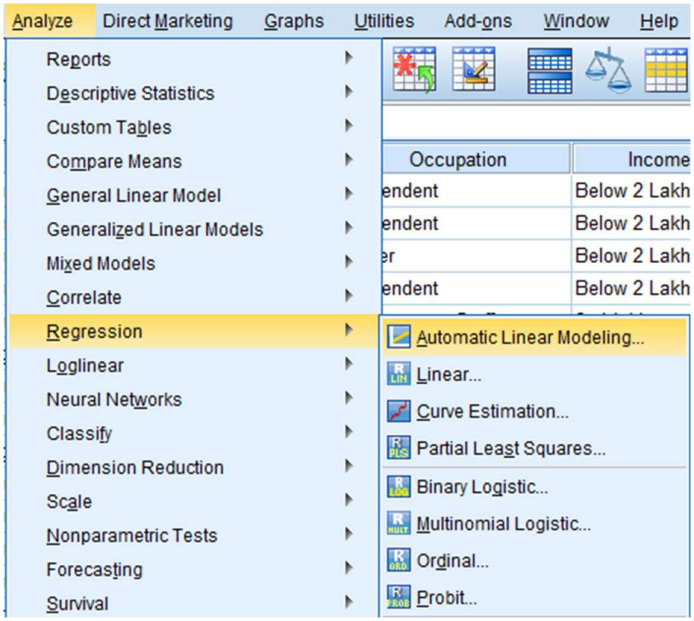
Field requirements (လိုအပ်သည့် အချက်အလက်များ)

Target တစ်ခုနဲ့ Input အနည်းဆုံး တစ်ခုတော့ ရှိရပါမယ်။ Target က continuous (scale level) ဖြစ်ရပါမယ်။ Predictors (Inputs) တွေအတွက်တော့ measurement level ကန့်သတ်ချက် မရှိပါဘူး။ Categorical (nominal နဲ့ ordinal) fields တွေကို factors တွေအနေနဲ့ သုံးပြီး၊ Continuous fields တွေကိုတော့ covariates တွေအနေနဲ့ သုံးပါတယ်။

မှတ်ချက်- တကယ်လို့ Categorical field တစ်ခုမှာ categories အရေအတွက် ၁၀၀၀ ထက်ကျော်နေရင်တော့ procedure က အလုပ်လုပ်မှာ မဟုတ်သလို model လည်း တည်ဆောက်ပေးမှာ မဟုတ်ပါဘူး။

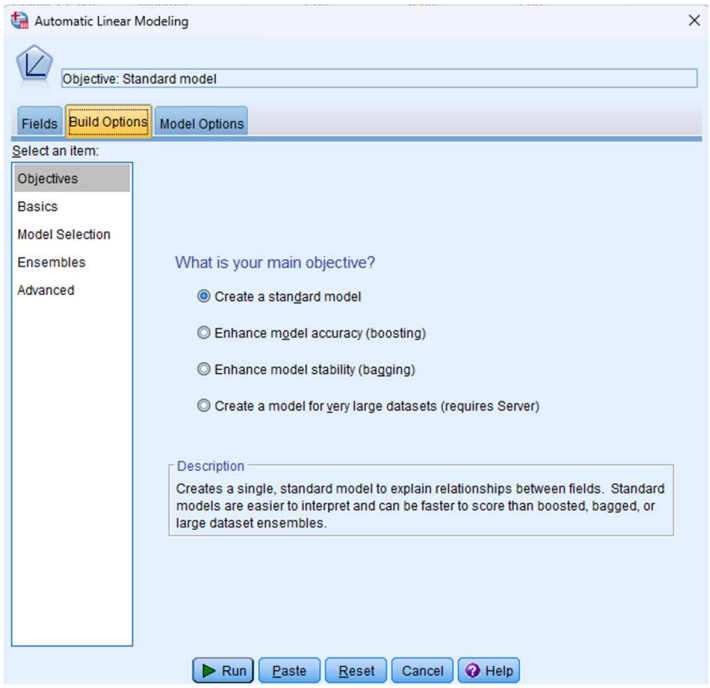
To obtain a linear model (Linear model ရယူအသုံးပြုရန် အဆင့်များ)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Regression > Automatic Linear Models...

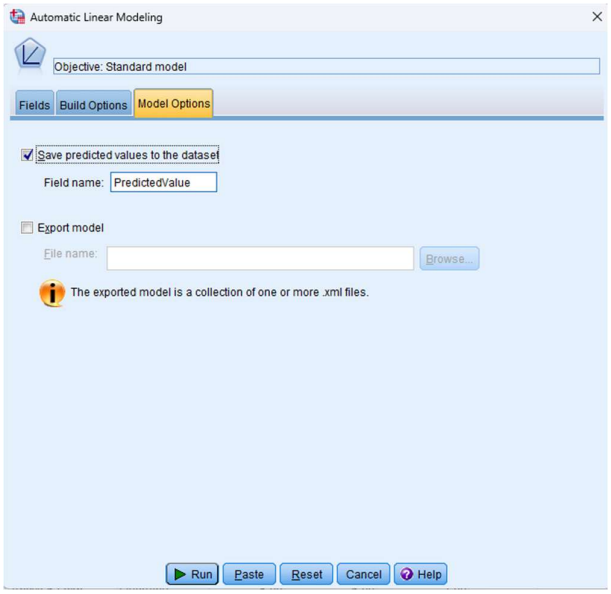


၂။ အနည်းဆုံး target တစ်ခုနဲ့ input တစ်ခု ရှိနေပါစေ။

၃။ Model တည်ဆောက်မှုဆိုင်ရာ အပြင်အဆင်တွေ သတ်မှတ်ဖို့ **Build Options** ကို နှိပ်ပါ။



၄။ ရမှတ် (scores) တွေကို dataset ထဲ save ဖို့နဲ့ model ကို ပြင်ပဖိုင်အနေနဲ့ export လုပ်ဖို့ **Model Options** ကို နှိပ်ပါ။



၅။ Procedure ကို run ဖို့နဲ့ Model တွေ ဖန်တီးဖို့ **Run** ကို နှိပ်ပါ။

Objectives (ရည်ရွယ်ချက်များ)

သင့်ရဲ့ အဓိက ရည်ရွယ်ချက်က ဘာလဲဆိုတာကို ရွေးချယ်ပေးရပါမယ်-

- **Create a standard model:** Predictors တွေကို သုံးပြီး target ကို ခန့်မှန်းဖို့ standard model တစ်ခုကို တည်ဆောက်တာပါ။ အများအားဖြင့် standard models တွေက ရှင်းပြရလွယ်ကူပြီး၊ ensemble models တွေထက် ပိုမြန်ဆန်ပါတယ်။
- **Enhance model accuracy (boosting):** ပိုမိုတိကျတဲ့ ခန့်မှန်းချက်တွေရဖို့ မော်ဒယ်တွေကို အဆင့်ဆင့် ဆက်တိုက်ဖန်တီးပေးတဲ့ boosting နည်းလမ်းကိုသုံးပြီး ensemble model တစ်ခု တည်ဆောက်တာပါ။ တည်ဆောက်ဖို့ အချိန်ပိုယူရနိုင်ပါတယ်။
- **Enhance model stability (bagging):** ပိုမိုယုံကြည်စိတ်ချရတဲ့ ခန့်မှန်းချက်တွေရဖို့ မော်ဒယ်အများကြီးကို ဖန်တီးပေးတဲ့ bagging (bootstrap aggregating) နည်းလမ်းကိုသုံးပြီး ensemble model တည်ဆောက်တာပါ။
- **Create a model for very large datasets:** အရမ်းကြီးမားတဲ့ dataset တွေအတွက် data တွေကို blocks လေးတွေခွဲပြီး ensemble model တည်ဆောက်တာပါ။ (ဒီ option က IBM SPSS Statistics Server လိုအပ်ပါတယ်) ။

Basics (အခြေခံများ)

Automatically prepare data (အလိုအလျောက် Data ပြင်ဆင်ခြင်း) ဒီ option က model ရဲ့ ခန့်မှန်းနိုင်စွမ်း (predictive power) ကို အကောင်းဆုံးဖြစ်စေဖို့အတွက် target နဲ့ predictors တွေကို အတွင်းပိုင်းမှာ အလိုအလျောက် ပြောင်းလဲ (transform) ပေးပါတယ်။ ပြောင်းလဲမှု မှန်သမျှကို model နဲ့အတူ save လုပ်ထားပြီး နောက်ထပ် scoring လုပ်မယ့် data အသစ်တွေအတွက် အသုံးပြုပေးမှာပါ။ Transform လုပ်ပြီးသား fields တွေရဲ့ မူရင်း (original) အချက်အလက်တွေကိုတော့ model ထဲကနေ ဖယ်ထုတ်ထားမှာ ဖြစ်ပါတယ်။ ပုံမှန်အားဖြင့် (By default) အောက်ပါ အလိုအလျောက် data ပြင်ဆင်မှုတွေကို လုပ်ဆောင်ပေးပါတယ်-

- **Date and Time handling (ရက်စွဲနှင့် အချိန် ကိုင်တွယ်ခြင်း)** ။ Date predictor (ရက်စွဲ ခန့်မှန်းကိန်း) တစ်ခုစီကို reference date (1970-01-01) ကနေစပြီး ကြာမြင့်ခဲ့တဲ့အချိန်တွေ ပါဝင်တဲ့ continuous predictor အသစ်တစ်ခုအဖြစ် ပြောင်းလဲပေးပါတယ်။ Time predictor (အချိန် ခန့်မှန်းကိန်း) တစ်ခုစီကိုလည်း reference time (00:00:00) ကနေစပြီး ကြာမြင့်ခဲ့တဲ့အချိန်တွေ ပါဝင်တဲ့ continuous predictor အသစ်တစ်ခုအနေနဲ့ ပြောင်းလဲပေးပါတယ်။

- **Adjust measurement level (တိုင်းတာမှုအဆင့် ချိန်ညှိခြင်း)** ။ ကွဲပြားတဲ့ တန်ဖိုး (distinct values) ၅ ခုထက် နည်းတဲ့ continuous predictors တွေကို ordinal predictors တွေအဖြစ် ပြောင်းလဲသတ်မှတ် (recast) ပါတယ်။ ကွဲပြားတဲ့ တန်ဖိုး ၁၀ ခုထက် များနေတဲ့ ordinal predictors တွေကိုတော့ continuous predictors တွေအဖြစ် ပြန်လည်ပြောင်းလဲ သတ်မှတ်ပေးပါတယ်။

- **Outlier handling (လွန်ကဲတန်ဖိုးများကို ကိုင်တွယ်ဖြေရှင်းခြင်း)** ။ Mean ကနေ 3 standard deviations အကွာအဝေး (cutoff value) ထက်ကျော်လွန်နေတဲ့ continuous predictors တွေရဲ့ တန်ဖိုးတွေကို အဲဒီ cutoff value အဖြစ်ပဲ သတ်မှတ်ပေးလိုက်ပါတယ်။

- **Missing value handling (ပျောက်ဆုံးတန်ဖိုးများကို ကိုင်တွယ်ဖြေရှင်းခြင်း)** ။ Nominal predictors တွေရဲ့ missing values တွေကို training partition ရဲ့ mode (အကြိမ်အများဆုံးတန်ဖိုး) နဲ့ အစားထိုးပေးပါတယ်။ Ordinal predictors တွေရဲ့ missing values တွေကိုတော့ training partition ရဲ့ median (အလယ်ကိန်း) နဲ့ အစားထိုးပေးပြီး၊ Continuous predictors တွေရဲ့ missing values တွေကိုတော့ training partition ရဲ့ mean (ပျမ်းမျှကိန်း) နဲ့ အစားထိုးပေးပါတယ်။

- **Supervised merging (ကြီးကြပ်မှုဖြင့် ပေါင်းစည်းခြင်း)**။ ဒါက target နဲ့ ဆက်စပ်ပြီး process လုပ်ရမယ့် fields အရေအတွက်ကို လျှော့ချပေးခြင်းအားဖြင့် model ကို ပိုပြီး ရိုးရှင်းကျစ်လျစ်စေပါတယ်။ Input နဲ့ target ကြားက ဆက်စပ်မှုအပေါ် မူတည်ပြီး ဆင်တူတဲ့ categories တွေကို ခွဲခြားသတ်မှတ်ပါတယ်။ သိသိသာသာ ကွာခြားမှုမရှိတဲ့ (ဆိုလိုတာက p-

value ဟာ 0.1 ထက် ကြီးနေတဲ့) categories တွေကို ပေါင်းစည်း (merge) လိုက်ပါတယ်။ တကယ်လို့ categories တွေအားလုံး တစ်ခုတည်းအဖြစ် ပေါင်းစည်းသွားမယ်ဆိုရင်တော့၊ သူတို့က predictor တစ်ခုအနေနဲ့ ဘာတန်ဖိုးမှ မရှိတော့တဲ့အတွက် အဲဒီ field ရဲ့ မူရင်း (original) ရော၊ ဆင်းသက်လာတဲ့ (derived) versions တွေကိုပါ model ထဲကနေ ဖယ်ထုတ်လိုက်မှာ ဖြစ်ပါတယ်။

- **Confidence level:** Coefficients တွေအတွက် ယုံကြည်မှု အတိုင်းအတာ ဖြစ်ပါတယ်။ 0 နဲ့ 100 ကြား သတ်မှတ်နိုင်ပြီး default က 95 ဖြစ်ပါတယ်။

Model Selection (မော်ဒယ် ရွေးချယ်ခြင်း)

Model selection method (မော်ဒယ်ရွေးချယ်မည့် နည်းလမ်း)

အောက်မှာ အသေးစိတ်ဖော်ပြထားတဲ့ မော်ဒယ်ရွေးချယ်မယ့် နည်းလမ်းတွေထဲက တစ်ခုကို ရွေးချယ်နိုင်သလို၊ ဒါမှမဟုတ် ရနိုင်သမျှ predictors အားလုံးကို main effects model terms တွေအနေနဲ့ ရိုးရိုးရှင်းရှင်းပဲ ထည့်သွင်းပေးမယ့် **Include all predictors** ကိုလည်း ရွေးချယ်နိုင်ပါတယ်။ ပုံမှန် (By default) အားဖြင့်တော့ Forward stepwise ကို အသုံးပြုထားပါတယ်။

Forward Stepwise Selection (ရှေ့သို့ တစ်ဆင့်ချင်း ရွေးချယ်ခြင်း)

ဒီနည်းလမ်းက မော်ဒယ်ထဲမှာ ဘာ effect မှ မပါဝင်ဘဲ စတင်ပါတယ်။ ပြီးတော့ သတ်မှတ်ထားတဲ့ stepwise criteria (စံသတ်မှတ်ချက်များ) အရ ထပ်ထည့်လို့ ဒါမှမဟုတ် ဖယ်ရှားလို့ မရတော့တဲ့အချိန်အထိ effects တွေကို တစ်ဆင့်ချင်းစီ ထည့်သွင်းတာ၊ ဖယ်ရှားတာတွေ လုပ်ဆောင်ပေးပါတယ်။

- **Criteria for entry/removal (ထည့်သွင်းရန်/ဖယ်ရှားရန် စံသတ်မှတ်ချက်)** ။ ဒါက effect တစ်ခုကို မော်ဒယ်ထဲ ထည့်သင့်သလား၊ ဒါမှမဟုတ် ဖယ်ရှားသင့်သလား ဆိုတာကို ဆုံးဖြတ်ပေးဖို့ သုံးတဲ့ statistic ဖြစ်ပါတယ်။
    - **Information Criterion (AICC)** ဟာ မော်ဒယ်အပေါ်မူတည်ပြီး training set ရဲ့ ဖြစ်နိုင်ခြေ (likelihood) အပေါ် အခြေခံထားတာဖြစ်ပြီး၊ အရမ်းရှုပ်ထွေးလွန်းတဲ့ မော်ဒယ်တွေကို ဒဏ်ကြေးပေးဖို့ (penalize လုပ်ဖို့) ချိန်ညှိထားပါတယ်။
    - **F Statistics** ကတော့ မော်ဒယ်ရဲ့ အမှားအယွင်း (model error) တိုးတက်ကောင်းမွန်လာမှုအပေါ် statistical test လုပ်ခြင်းအပေါ် အခြေခံထားတာပါ။

- **Adjusted R-squared** ဟာ training set ရဲ့ အံဝင်ခွင်ကျဖြစ်မှု (fit) အပေါ် အခြေခံပြီး အရမ်းရှုပ်ထွေးလွန်းတဲ့ မော်ဒယ်တွေကို ဒဏ်ကြေးပေးဖို့ ချိန်ညှိထားတာ ဖြစ်ပါတယ်။
- **Overfit Prevention Criterion (ASE)** ကတော့ overfit (အလွန်အကျွံ အံဝင်ခွင်ကျဖြစ်မှု) ကာကွယ်ရေးအုပ်စု ရဲ့ fit (average squared error ဒါမှမဟုတ် ASE) အပေါ် အခြေခံထားတာပါ။ ဒီ overfit prevention set ဆိုတာက မော်ဒယ်ကို train လုပ်ရာမှာ အသုံးမပြုဘဲ ဖယ်ချန်ထားတဲ့ မူရင်း dataset ရဲ့ ခန့်မှန်းခြေ ၃၀% ခန့်ရှိတဲ့ random subsample (ကျပန်းခွဲထုတ်ထားသော အုပ်စုငယ်) တစ်ခု ဖြစ်ပါတယ်။

- တကယ်လို့ F Statistics ကလွဲပြီး တခြား criterion တစ်ခုခုကို ရွေးချယ်ခဲ့မယ်ဆိုရင်၊ အဆင့်တစ်ဆင့်ချင်းစီတိုင်းမှာ criterion ကို အများဆုံး အပေါင်းလက္ခဏာဘက် တိုးတက်စေတဲ့ effect ကို မော်ဒယ်ထဲ ထည့်သွင်းပေးမှာ ဖြစ်ပါတယ်။ မော်ဒယ်ထဲမှာ criterion ကို ကျဆင်းစေတဲ့ effects မှန်သမျှကိုတော့ ဖယ်ရှားပစ်မှာ ဖြစ်ပါတယ်။
- တကယ်လို့ F Statistics ကို criterion အဖြစ် ရွေးချယ်ခဲ့မယ်ဆိုရင်တော့၊ အဆင့်တစ်ဆင့်ချင်းစီတိုင်းမှာ **Include effects with p-values less than** လို့ သတ်မှတ်ထားတဲ့ threshold (ကန့်သတ်တန်ဖိုး) ထက်ငယ်ပြီး အငယ်ဆုံး p-value ရှိတဲ့ effect ကို မော်ဒယ်ထဲ ထည့်သွင်းပေးပါတယ်။ (ပုံမှန် default တန်ဖိုးကတော့ 0.05 ပါ)။ မော်ဒယ်ထဲမှာ **Remove effects with p-values greater than** ဆိုတဲ့ ကန့်သတ်တန်ဖိုးထက်ကြီးနေတဲ့ p-value ရှိတဲ့ effects တွေကိုတော့ ဖယ်ရှားပစ်ပါတယ်။ (ပုံမှန် default တန်ဖိုးက 0.10 ပါ)။
- **Customize maximum number of effects in the final model (နောက်ဆုံးမော်ဒယ်ရှိ အများဆုံး effects အရေအတွက်ကို စိတ်ကြိုက်သတ်မှတ်ခြင်း)** ။ ပုံမှန်အားဖြင့်တော့ ရနိုင်သမျှ effects အားလုံးကို မော်ဒယ်ထဲ ထည့်သွင်းနိုင်ပါတယ်။ ဒါမှမဟုတ် တကယ်လို့ stepwise algorithm ဟာ သတ်မှတ်ထားတဲ့ အများဆုံး effects အရေအတွက်နဲ့ အဆင့်တစ်ဆင့်ကို အဆုံးသတ်သွားပြီဆိုရင်၊ အဲဒီ algorithm ဟာ လက်ရှိ ရရှိထားတဲ့ effects တွေနဲ့တင် ဆက်မလုပ်တော့ဘဲ ရပ်တန့်သွားမှာ ဖြစ်ပါတယ်။
- **Customize maximum number of steps (အများဆုံး အဆင့်အရေအတွက်ကို စိတ်ကြိုက်သတ်မှတ်ခြင်း)** ။ Stepwise algorithm ဟာ သတ်မှတ်ထားတဲ့ အဆင့်အရေအတွက်တစ်ခု ရောက်ရင် ရပ်တန့်သွားပါတယ်။ ပုံမှန်အားဖြင့် ဒါဟာ ရရှိနိုင်တဲ့ effects အရေအတွက်ရဲ့ ၃ ဆ ဖြစ်ပါတယ်။ အဲဒီလိုမဟုတ်ဘဲ သင်ကိုယ်တိုင် အများဆုံး အဆင့်အရေအတွက်အတွက် positive integer (အပေါင်းကိန်းပြည့်) တစ်ခုကို သတ်မှတ်ပေးလို့လည်း ရပါတယ်။

Best Subsets Selection (အကောင်းဆုံး အုပ်စုငယ်များ ရွေးချယ်ခြင်း)

ဒီနည်းလမ်းက best subsets criterion အပေါ်မူတည်ပြီး အကောင်းဆုံးကို ရွေးချယ်နိုင်ဖို့အတွက်၊ ဖြစ်နိုင်သမျှ မော်ဒယ် "အားလုံး" ဒါမှမဟုတ် အနည်းဆုံးတော့ forward stepwise ထက်စာရင် ပိုများတဲ့ ဖြစ်နိုင်ချေရှိတဲ့ မော်ဒယ်အုပ်စုတွေကို စစ်ဆေးပေးပါတယ်။ Information Criterion (AICC) ဟာ training set ရဲ့ likelihood အပေါ် အခြေခံပြီး အရမ်းရှုပ်ထွေးတဲ့ မော်ဒယ်ကို ဒဏ်ပေးဖို့ ချိန်ညှိပါတယ်။ Adjusted R-squared ဟာ training set ရဲ့ fit အပေါ် အခြေခံပြီး အရမ်းရှုပ်ထွေးတဲ့ မော်ဒယ်ကို ဒဏ်ပေးဖို့ ချိန်ညှိပါတယ်။ Overfit Prevention Criterion (ASE) ကတော့ မော်ဒယ်ကို train ရာမှာ မသုံးဘဲ ဖယ်ချန်ထားတဲ့ မူရင်း dataset ရဲ့ ၃၀% ခန့်ရှိတဲ့ overfit prevention set ရဲ့ fit အပေါ် အခြေခံပါတယ်။ နောက်ဆုံးမှာတော့ Criterion ရဲ့ တန်ဖိုးအကြီးဆုံးရရှိတဲ့ မော်ဒယ်ကို အကောင်းဆုံး မော်ဒယ်အဖြစ် ရွေးချယ်ပါတယ်။

မှတ်ချက်- Best subsets selection က forward stepwise selection ထက်စာရင် ကွန်ပျူတာတွက်ချက်မှုပိုင်း (computational) မှာ ပိုပြီး အချိန်နဲ့ အင်အား ယူရပါတယ်။ Best subsets ကို boosting, bagging ဒါမှမဟုတ် အရမ်းကြီးမားတဲ့ datasets တွေနဲ့ တွဲသုံးတဲ့အခါ၊ forward stepwise selection သုံးပြီး တည်ဆောက်တဲ့ standard model ထက်စာရင် အချိန်သိသိသာသာ ပိုကြာနိုင်ပါတယ်။

Ensembles (အစုအဖွဲ့များ / အစုလိုက်မော်ဒယ်များ)

ဒီ settings တွေက Objectives ထဲမှာ boosting, bagging, ဒါမှမဟုတ် အရမ်းကြီးမားတဲ့ datasets (very large datasets) တွေကို ရွေးချယ်တောင်းဆိုထားတဲ့အခါ အလုပ်လုပ်မယ့် ensembling ရဲ့ အမူအကျင့် (behavior) ကို သတ်မှတ်ပေးပါတယ်။ ကိုယ်ရွေးချယ်ထားတဲ့ objective နဲ့ မသက်ဆိုင်တဲ့ options တွေကိုတော့ လျစ်လျူရှု (ignore) ထားမှာပါ။

- **Bagging and Very Large Datasets (Bagging နည်းလမ်းနှင့် အလွန်ကြီးမားသော ဒေတာအစုအဝေးများ) ။** Ensemble ကို အမှတ်ပေး (scoring) တွက်ချက်တဲ့အခါ၊ ဒါဟာ အခြေခံမော်ဒယ် (base models) တွေဆီကနေ ရလာတဲ့ ခန့်မှန်းတန်ဖိုး (predicted values) တွေကို ပေါင်းစပ်ပြီး ensemble score value တွက်ထုတ်ဖို့ သုံးတဲ့ စည်းမျဉ်းတစ်ခု ဖြစ်ပါတယ်။

- **Default combining rule for continuous targets (ဆက်တိုက်ပြောင်းလဲနေသော targets များအတွက် ပုံမှန်ပေါင်းစပ်မှုစည်းမျဉ်း)။** Continuous targets တွေအတွက် ensemble ကနေ ခန့်မှန်းပေးတဲ့တန်ဖိုးတွေကို အခြေခံမော်ဒယ်တွေကနေရလာတဲ့ ခန့်မှန်းတန်ဖိုးတွေရဲ့ mean (ပျမ်းမျှကိန်း) ဒါမှမဟုတ် median (အလယ်ကိန်း) တွေကို သုံးပြီး ပေါင်းစပ်နိုင်ပါတယ်။ *မှတ်သားထားရမယ့်အချက်က-* မော်ဒယ်ရဲ့ တိကျမှုကို

ပိုကောင်းလာစေဖို့ (enhance model accuracy) ကိုသာ ရည်ရွယ်ချက် (objective) အနေနဲ့ ထားတဲ့အခါ၊ ပေါင်းစပ်ဖို့ ရွေးချယ်ထားတဲ့ စည်းမျဉ်း (combining rule selections) တွေကို လျစ်လျူရှုထားမှာ ဖြစ်ပါတယ်။ Boosting က categorical targets တွေကို အမှတ်ပေးတွက်ချက်ဖို့ အမြဲတမ်း weighted majority vote (အလေးချိန်ပေးထားသော အများစုမဲ) ကို သုံးပြီး၊ continuous targets တွေအတွက်ဆိုရင်တော့ weighted median ကို အသုံးပြုပါတယ်။

- **Boosting and Bagging**။ မော်ဒယ်ရဲ့ တိကျမှု (accuracy) သို့မဟုတ် တည်ငြိမ်မှု (stability) ကို ပိုကောင်းလာစေဖို့ ရည်ရွယ်တဲ့အခါ တည်ဆောက်မယ့် အခြေခံမော်ဒယ် (base models) အရေအတွက်ကို သတ်မှတ်ပေးပါ။ Bagging အတွက်ဆိုရင် ဒါဟာ bootstrap samples အရေအတွက် ဖြစ်ပါတယ်။ ဂဏန်းတန်ဖိုးဟာ အပေါင်းကိန်းပြည့် (positive integer) ဖြစ်ရပါမယ်။

Advanced (အဆင့်မြင့် ရွေးချယ်စရာများ)

- **Replicate results (ရလဒ်များကို ထပ်တူပြန်ထုတ်ယူခြင်း)**။ Random seed (ကျပန်းကိန်းဂဏန်းအစ) ကို သတ်မှတ်ပေးခြင်းက သင့်ရဲ့ analysis တွေကို အတိအကျ ထပ်တူပြန်လည်ရရှိအောင် (replicate) လုပ်ဆောင်နိုင်စေပါတယ်။ Overfit prevention set ထဲမှာ ဘယ် records တွေ ပါဝင်မလဲဆိုတာကို ရွေးချယ်ဖို့ random number generator (ကျပန်းကိန်းဂဏန်း ထုတ်ပေးသည့်စနစ်) ကို အသုံးပြုပါတယ်။ ကိန်းပြည့်တစ်ခုကို ကိုယ်တိုင်ရိုက်ထည့်နိုင်သလို၊ 1 နဲ့ 2147483647 ကြားမှာရှိတဲ့ pseudo-random integer တစ်ခုကို အလိုအလျောက် ဖန်တီးပေးဖို့ **Generate** ကိုလည်း နှိပ်နိုင်ပါတယ်။ ပုံမှန် (default) တန်ဖိုးကတော့ 54752075 ဖြစ်ပါတယ်။

Model Options (မော်ဒယ်ဆိုင်ရာ ရွေးချယ်စရာများ)

- **Save predicted values to the dataset (ခန့်မှန်းတန်ဖိုးများကို dataset ထဲသို့ သိမ်းဆည်းခြင်း)** ။ ပုံမှန် variable နာမည် (default variable name) ကတော့ Predicted Value ဖြစ်ပါတယ်။

- **Export model (မော်ဒယ်ကို အပြင်သို့ ထုတ်ယူခြင်း)** ။ ဒါက မော်ဒယ်ကို ပြင်ပ .zip ဖိုင်တစ်ခုအနေနဲ့ ရေးသားသိမ်းဆည်းပေးတာပါ။ ဒီမော်ဒယ်ဖိုင်ကို အခြား data ဖိုင်တွေမှာ အမှတ်ပေးခန့်မှန်းဖို့ (scoring purposes) အတွက် မော်ဒယ်အချက်အလက်တွေ အသုံးချရာမှာ ပြန်လည်အသုံးပြုနိုင်ပါတယ်။ ထူးခြားပြီး မှန်ကန်တဲ့ ဖိုင်နာမည် (unique, valid file name) တစ်ခုကို သတ်မှတ်ပေးပါ။ တကယ်လို့ သတ်မှတ်လိုက်တဲ့ ဖိုင်နာမည်က

ရှိပြီးသားဖိုင်တစ်ခုခုနဲ့ တူနေမယ်ဆိုရင်တော့ အဲဒီဖိုင်အဟောင်းပေါ်မှာ ထပ်မံရေးသား (overwrite) သွားမှာ ဖြစ်ပါတယ်။

Model Summary (မော်ဒယ် အကျဉ်းချုပ်)

Model Summary မြင်ကွင်းဟာ မော်ဒယ်နဲ့ သူ့ရဲ့ အံဝင်ခွင်ကျဖြစ်မှု (fit) ကို တစ်ချက်ကြည့်ရုံနဲ့ သိနိုင်အောင် အကျဉ်းချုပ် (snapshot) ဖော်ပြပေးထားတာ ဖြစ်ပါတယ်။

- **Table (ဇယား)။** ဇယားထဲမှာ အောက်ပါ အချက်တွေအပါအဝင် အဆင့်မြင့် မော်ဒယ် settings (high-level model settings) တချို့ကို ဖော်ပြပေးထားပါတယ်-
    - **Fields** tab မှာ သတ်မှတ်ထားတဲ့ target ရဲ့ နာမည်၊
    - **Basics** settings မှာ သတ်မှတ်ထားတဲ့အတိုင်း အလိုအလျောက် data ပြင်ဆင်မှု (automatic data preparation) တွေကို လုပ်ဆောင်ခဲ့ခြင်း ရှိမရှိ၊
    - **Model Selection** settings မှာ သတ်မှတ်ထားတဲ့ မော်ဒယ်ရွေးချယ်တဲ့ နည်းလမ်းနဲ့ ရွေးချယ်မှု စံသတ်မှတ်ချက် (selection criterion) ။ နောက်ဆုံးရရှိလာတဲ့ မော်ဒယ်အတွက် selection criterion ရဲ့ တန်ဖိုးကိုလည်း ပြသပေးထားပြီး၊ ၎င်းကို 'တန်ဖိုးငယ်လေ ပိုကောင်းလေ' (smaller is better) ဆိုတဲ့ ပုံစံနဲ့ ဖော်ပြပေးထားပါတယ်။
- **Chart (ပုံပြဇယား)။** Chart က နောက်ဆုံးရရှိလာတဲ့ မော်ဒယ်ရဲ့ တိကျမှန်ကန်မှု (accuracy) ကို ပြသပေးပြီး၊ သူ့ကိုတော့ 'တန်ဖိုးကြီးလေ ပိုကောင်းလေ' (larger is better) ပုံစံနဲ့ ပြသထားပါတယ်။ အဲဒီတန်ဖိုးဟာ နောက်ဆုံးမော်ဒယ်အတွက် adjusted R^2 ကို 100 နဲ့ မြှောက်ထားတဲ့ တန်ဖိုးပဲ ဖြစ်ပါတယ်။

Automatic Data Preparation (အလိုအလျောက် Data ပြင်ဆင်ခြင်း)

ဒီမြင်ကွင်း (view) မှာ automatic data preparation (ADP) အဆင့်မှာ ဘယ် fields တွေကို ဖယ်ထုတ်ခဲ့တယ်၊ ပြီးတော့ ပြောင်းလဲထားတဲ့ (transformed) fields တွေကို ဘယ်လို ရယူခဲ့တယ်ဆိုတဲ့ အချက်အလက်တွေကို ပြသပေးပါတယ်။ ပြောင်းလဲခဲ့တဲ့ ဒါမှမဟုတ် ဖယ်ထုတ်ခဲ့တဲ့ field တစ်ခုချင်းစီအတွက်၊ ဇယားထဲမှာ field နာမည်၊ analysis မှာ သူပါဝင်တဲ့ အခန်းကဏ္ဍနဲ့ ADP အဆင့်မှာ လုပ်ဆောင်ခဲ့တဲ့ အရေးယူဆောင်ရွက်ချက် (action) တွေကို စာရင်းပြုစုထားပါတယ်။ Fields တွေကို သူတို့ရဲ့ နာမည် အကွာရာစဉ်အလိုက် (အငယ်မှအကြီး) စီစဉ်ထားပါတယ်။ Field တစ်ခုစီအတွက် လုပ်ဆောင်နိုင်တဲ့ actions တွေကတော့ အောက်ပါအတိုင်း ဖြစ်ပါတယ်-

- **Derive duration: months** က ရက်စွဲ (dates) တွေ ပါဝင်တဲ့ field တစ်ခုထဲက တန်ဖိုးတွေကနေ လက်ရှိ စနစ်ရဲ့ ရက်စွဲ (current system date) အထိ ကြာမြင့်ခဲ့တဲ့ အချိန်ကို လ (months) နဲ့ တွက်ချက်ပေးပါတယ်။
- **Derive duration: hours** က အချိန် (times) တွေ ပါဝင်တဲ့ field တစ်ခုထဲက တန်ဖိုးတွေကနေ လက်ရှိ စနစ်ရဲ့ အချိန် (current system time) အထိ ကြာမြင့်ခဲ့တဲ့ အချိန်ကို နာရီ (hours) နဲ့ တွက်ချက်ပေးပါတယ်။
- **Change measurement level from continuous to ordinal** က ကွဲပြားတဲ့တန်ဖိုး (unique values) ၅ ခုထက် နည်းတဲ့ continuous fields တွေကို ordinal fields တွေအဖြစ် ပြောင်းလဲသတ်မှတ် (recast) ပေးပါတယ်။
- **Change measurement level from ordinal to continuous** ကတော့ ကွဲပြားတဲ့တန်ဖိုး ၁၀ ခုထက် များတဲ့ ordinal fields တွေကို continuous fields တွေအဖြစ် ပြောင်းလဲသတ်မှတ်ပေးပါတယ်။
- **Trim outliers** က cutoff value (mean ကနေ 3 standard deviations အကွာအဝေး) ထက် ကျော်လွန်နေတဲ့ continuous predictors တွေရဲ့ တန်ဖိုးတွေကို အဲဒီ cutoff value အဖြစ်ပဲ သတ်မှတ်ပေးလိုက်ပါတယ်။
- **Replace missing values** က nominal fields တွေရဲ့ missing values (ပျောက်ဆုံးတန်ဖိုးများ) ကို mode (အကြိမ်အများဆုံးတန်ဖိုး) နဲ့ အစားထိုးပေးပြီး၊ ordinal fields တွေအတွက်ဆိုရင် median (အလယ်ကိန်း) နဲ့ အစားထိုးပေးကာ၊ continuous fields တွေအတွက်ဆိုရင်တော့ mean (ပျမ်းမျှကိန်း) နဲ့ အစားထိုးပေးပါတယ်။
- **Merge categories to maximize association with target** က input နဲ့ target ကြားက ဆက်စပ်မှုအပေါ် မူတည်ပြီး "ဆင်တူတဲ့" predictor categories တွေကို ခွဲခြားသတ်မှတ်ပေးပါတယ်။ သိသိသာသာ ကွာခြားမှုမရှိတဲ့ (ဆိုလိုတာက p-value ဟာ 0.05 ထက် ကြီးနေတဲ့) categories တွေကို ပေါင်းစည်း (merge) လိုက်ပါတယ်။
- **Exclude constant predictor / after outlier handling / after merging of categories** က တခြား ADP လုပ်ဆောင်ချက်တွေ လုပ်ပြီးသွားတဲ့အချိန်မှာ တန်ဖိုးတစ်ခုတည်း (single value) ပဲ ကျန်ရှိနေတော့တဲ့ predictors တွေကို ဖယ်ရှားပစ်ပါတယ်။

Predictor Importance (ခန့်မှန်းကိန်း၏ အရေးပါမှု)

ပုံမှန်အားဖြင့်တော့ သင်က မော်ဒယ်တည်ဆောက်တဲ့အခါ အရေးအပါဆုံးဖြစ်တဲ့ predictor fields တွေအပေါ်မှာပဲ အာရုံစိုက်ချင်မှာဖြစ်ပြီး၊ အရေးမပါလှတဲ့ အရာတွေကို ချန်ထားခဲ့ဖို့ ဒါမှမဟုတ်

လျစ်လျူရှုဖို့ စဉ်းစားပါလိမ့်မယ်။ Predictor importance chart (ခန့်မှန်းကိန်း အရေးပါမှုပြ ဇယား) က မော်ဒယ်ကို ခန့်မှန်းတွက်ချက်ရာမှာ predictor တစ်ခုချင်းစီရဲ့ နှိုင်းရအရေးပါမှု (relative importance) ကို ညွှန်ပြပေးခြင်းအားဖြင့် သင့်ကို ကူညီပေးပါတယ်။ တန်ဖိုးတွေက နှိုင်းရတန်ဖိုးတွေ ဖြစ်တဲ့အတွက် ပြသထားတဲ့ predictors အားလုံးရဲ့ တန်ဖိုးတွေ ပေါင်းလိုက်ရင် 1.0 ရရှိမှာ ဖြစ်ပါတယ်။ Predictor ရဲ့ အရေးပါမှုဆိုတာ မော်ဒယ်ရဲ့ တိကျမှန်ကန်မှု (model accuracy) နဲ့ သက်ဆိုင်တာ မဟုတ်ပါဘူး။ သူက ခန့်မှန်းမှုတစ်ခု လုပ်ဆောင်ရာမှာ predictor တစ်ခုချင်းစီက ဘယ်လောက် အရေးပါသလဲ ဆိုတာကိုပဲ ရည်ညွှန်းတာဖြစ်ပြီး၊ အဲဒီခန့်မှန်းချက်က တိကျမှန်ကန်မှု ရှိမရှိ ဆိုတာကို ဆိုလိုတာ မဟုတ်ပါဘူး။

Predicted By Observed (လေ့လာတွေ့ရှိချက်အပေါ်မူတည်၍ ခန့်မှန်းချက်)

ဒါက vertical axis (ဒေါင်လိုက်ဝင်ရိုး) မှာ predicted values (ခန့်မှန်းတန်ဖိုးများ) နဲ့ horizontal axis (အလျားလိုက်ဝင်ရိုး) မှာ observed values (လေ့လာတွေ့ရှိရတဲ့ တန်ဖိုးများ) ကိုထားပြီး binned scatterplot တစ်ခုကို ပြသပေးပါတယ်။ အကောင်းဆုံးကတော့ အစက်လေးတွေဟာ ၄၅ ဒီဂရီ မျဉ်းပေါ်မှာ ကျရောက်နေသင့်ပါတယ်၊ ဒီမြင်ကွင်းက မော်ဒယ်ကနေ အလွန်ဆိုးရွားစွာ ခန့်မှန်းထားတဲ့ မှတ်တမ်းတွေ (records) ရှိနေသလားဆိုတာကို သင့်ကို ပြောပြပေးနိုင်ပါတယ်။

Residuals (အကြွင်းအကျန်များ / ကွာဟချက်များ)

ဒါက မော်ဒယ်ရဲ့ residuals (ကွာဟချက်များ) နဲ့ပတ်သက်တဲ့ diagnostic chart (ရောဂါရှာဖွေရေး/စစ်ဆေးရေး ပုံပြဇယား) တစ်ခုကို ပြသပေးပါတယ်။

Chart styles (ဇယားပုံစံများ)

Style dropdown list ကနေ ရွေးချယ်ဝင်ရောက်နိုင်တဲ့ မတူညီတဲ့ ပြသမှုပုံစံတွေ ရှိပါတယ်။

- **Histogram (ဟစ္စတိုဂရမ်)။** ဒါက normal distribution (ပုံမှန်ဖြန့်ကျက်မှု) ကို ထပ်မံထည့်သွင်းဖော်ပြထားတဲ့ studentized residuals တွေရဲ့ binned histogram တစ်ခု ဖြစ်ပါတယ်။ Linear models တွေက residuals တွေဟာ normal distribution ရှိတယ်လို့ ယူဆထားတဲ့အတွက်၊ အကောင်းဆုံးကတော့ ဒီ histogram ဟာ ချောမွေ့တဲ့မျဉ်း (smooth line) နဲ့ အနီးစပ်ဆုံး တူညီနေသင့်ပါတယ်။
- **P-P Plot (P-P ပုံပြဇယား)။** ဒါက studentized residuals တွေကို normal distribution နဲ့ နှိုင်းယှဉ်ပြတဲ့ binned probability-probability plot တစ်ခု ဖြစ်ပါတယ်။ အကယ်၍ plot လုပ်ထားတဲ့ အစက်လေးတွေရဲ့ ဆင်ခြေလျှော (slope) ဟာ ပုံမှန်မျဉ်း (normal line) ထက် မတ်စောက်မှုနည်းနေရင်၊ residuals တွေဟာ normal distribution ထက်ပိုပြီး ပြောင်းလဲနိုင်စွမ်း (variability) ကြီးမားနေတာကို ပြသပါတယ်။ အကယ်၍ ဆင်ခြေလျှောက

ပိုပြီး မတ်စောက်နေရင်တော့၊ residuals တွေဟာ normal distribution ထက် variability ပိုနည်းနေတာကို ပြသပါတယ်။ အကယ်၍ plot လုပ်ထားတဲ့ အစက်လေးတွေက S ပုံသဏ္ဌာန် ကွေးနေမယ်ဆိုရင်တော့၊ residuals တွေရဲ့ ဖြန့်ကျက်မှုက တိမ်းစောင်း (skewed) နေပါတယ်။

Outliers (လွန်ကဲတန်ဖိုးများ)

ဒီဇယားက မော်ဒယ်အပေါ်မှာ မလိုလားအပ်တဲ့ လွှမ်းမိုးမှု (undue influence) ရှိနေတဲ့ records (မှတ်တမ်းများ) ကို စာရင်းပြုစုပေးပြီး၊ (Fields tab မှာ သတ်မှတ်ထားရင်) record ID၊ target value (ပစ်မှတ်တန်ဖိုး) နဲ့ Cook's distance တို့ကို ပြသပေးပါတယ်။ Cook's distance ဆိုတာ တကယ်လို့ မော်ဒယ်မြောက်ဖော်ကိန်းများ (model coefficients) တွက်ချက်မှုထဲကနေ သီးခြား record တစ်ခုကို ဖယ်ထုတ်လိုက်မယ်ဆိုရင်၊ records အားလုံးရဲ့ residuals တွေ ဘယ်လောက်အထိ ပြောင်းလဲသွားမလဲဆိုတာကို တိုင်းတာတဲ့ အတိုင်းအတာပါ။ ကြီးမားတဲ့ Cook's distance တန်ဖိုးရှိနေခြင်းက အဲဒီ record ကို ဖယ်ထုတ်လိုက်ရင် coefficients တွေကို သိသိသာသာ ပြောင်းလဲသွားစေတယ်ဆိုတာကို ညွှန်ပြနေတဲ့အတွက်၊ သူ့ကို လွှမ်းမိုးမှုရှိသူ (influential) အဖြစ် သတ်မှတ်စဉ်းစားသင့်ပါတယ်။

ဒီလို လွှမ်းမိုးမှုရှိတဲ့ records တွေကို မော်ဒယ်ခန့်မှန်းတွက်ချက်တဲ့အခါ အလေးချိန် (weight) လျှော့ပေးမလား၊ ဒါမှမဟုတ် လွန်ကဲနေတဲ့တန်ဖိုး (outlying values) တွေကို လက်ခံနိုင်တဲ့ ကန့်သတ်ချက် (acceptable threshold) တစ်ခုအထိ ဖြတ်ချ (truncate) မလား၊ ဒါမှမဟုတ် ဒီ influential records တွေကို လုံးဝ ဖယ်ရှားပစ်မလားဆိုတာ ဆုံးဖြတ်နိုင်ဖို့ သေချာ ဂရုတစိုက် စစ်ဆေးသင့်ပါတယ်။

Effects (သက်ရောက်မှုများ)

ဒီမြင်ကွင်းက မော်ဒယ်ထဲမှာရှိတဲ့ effect (သက်ရောက်မှု) တစ်ခုချင်းစီရဲ့ အရွယ်အစားကို ပြသပေးပါတယ်။

Styles (စတိုင်များ)

Style dropdown list ကနေ ရွေးချယ်ဝင်ရောက်နိုင်တဲ့ မတူညီတဲ့ ပြသမှုပုံစံတွေ ရှိပါတယ်။

- **Diagram (ပုံကြမ်း)။** ဒါက effects တွေကို predictor importance (ခန့်မှန်းကိန်း အရေးပါမှု) အကြီးမှအငယ် (decreasing) အတိုင်း အပေါ်ကနေအောက် စီစဉ်ပြသထားတဲ့ chart တစ်ခု ဖြစ်ပါတယ်။ ပုံထဲက ဆက်သွယ်ထားတဲ့ မျဉ်း (connecting lines) တွေကို effect significance (သက်ရောက်မှု အရေးပါမှု) အပေါ် မူတည်ပြီး အလေးချိန်ပေး (weighted) ထားပါတယ်၊ မျဉ်းအထူ (line width) ပိုကြီးလေ ပိုပြီး အရေးပါတဲ့ effects (ပိုသေးငယ်တဲ့ p-values) ကို ကိုယ်စားပြုလေ ဖြစ်ပါတယ်။ ဆက်သွယ်ထားတဲ့ မျဉ်းပေါ်ကို mouse တင်

(hover) လိုက်မယ်ဆိုရင် effect ရဲ့ p-value နဲ့ importance ကို ပြသပေးတဲ့ tooltip လေး ပေါ်လာမှာပါ။ ဒါက ပုံမှန် (default) အပြင်အဆင် ဖြစ်ပါတယ်။

- **Table (ဇယား)။** ဒါက မော်ဒယ်တစ်ခုလုံး (overall model) နဲ့ မော်ဒယ်ထဲက တစ်ခုချင်းစီရဲ့ effects တွေအတွက် ပြသပေးတဲ့ ANOVA table တစ်ခု ဖြစ်ပါတယ်။ တစ်ခုချင်းစီရဲ့ effects တွေကို predictor importance အကြီးမှအငယ်အတိုင်း အပေါ်ကနေအောက် စီစဉ်ထားပါတယ်။ သတိပြုရမယ့်အချက်က ပုံမှန် (by default) အားဖြင့် ဒီဇယားကို overall model ရဲ့ ရလဒ်တွေကိုပဲ ပြသဖို့ ခေါက်ထား (collapsed) ပေးထားပါတယ်။ မော်ဒယ်ရဲ့ သီးခြား effect တစ်ခုချင်းစီအတွက် ရလဒ်တွေကို ကြည့်ချင်တယ်ဆိုရင်တော့ ဇယားထဲက Corrected Model ဆဲလ် (cell) ကို နှိပ်ပါ။

Predictor importance (ခန့်မှန်းကိန်း အရေးပါမှု)

မြင်ကွင်းမှာ ဘယ် predictors တွေကို ပြသမလဲဆိုတာကို ထိန်းချုပ်ပေးတဲ့ Predictor Importance ဆလိုက်ဒါ (slider) တစ်ခု ရှိပါတယ်။ ဒီအရာက မော်ဒယ်ကို ပြောင်းလဲသွားစေတာ မဟုတ်ဘဲ၊ အရေးအပါဆုံး predictors တွေအပေါ်မှာပဲ သင့်ကို အာရုံစိုက်နိုင်စေဖို့ ကူညီပေးတာပါ။ ပုံမှန်အားဖြင့်တော့ ထိပ်ဆုံးက အရေးအပါဆုံး effects ၁၀ ခု (top 10 effects) ကို ပြသပေးထားပါတယ်။

Significance (အရေးပါမှု)

Predictor importance အပေါ်အခြေခံပြီး ပြသထားတာတွေအပြင်၊ ဘယ် effects တွေကို မြင်ကွင်းမှာ ဆက်လက်ပြသမလဲဆိုတာကို ထပ်မံထိန်းချုပ်ပေးတဲ့ Significance ဆလိုက်ဒါ (slider) တစ်ခု ရှိပါတယ်။ ဆလိုက်ဒါမှာ သတ်မှတ်ထားတဲ့တန်ဖိုးထက် ကြီးတဲ့ significance values ရှိတဲ့ effects တွေကို ဖျောက်ထား (hidden) ပေးပါလိမ့်မယ်။ ဒါကလည်း မော်ဒယ်ကို ပြောင်းလဲသွားစေတာ မဟုတ်ဘဲ၊ အရေးအပါဆုံး effects တွေအပေါ်မှာပဲ သင့်ကို အာရုံစိုက်နိုင်စေဖို့ ကူညီပေးတာပါ။ ပုံမှန်အားဖြင့် default တန်ဖိုးက 1.00 ဖြစ်တဲ့အတွက် significance အပေါ် မူတည်ပြီး ဘယ် effects ကိုမှ စစ်ထုတ် (filtered) ဖျောက်ထားမှာ မဟုတ်ပါဘူး။

Coefficients (မြှောက်ဖော်ကိန်းများ)

ဒီမြင်ကွင်းက မော်ဒယ်ထဲမှာရှိတဲ့ coefficient (မြှောက်ဖော်ကိန်း) တစ်ခုချင်းစီရဲ့ တန်ဖိုးကို ပြသပေးပါတယ်။ သတိပြုရမယ့်အချက်က factors (categorical predictors - အမျိုးအစားခွဲခြားထားတဲ့ ခန့်မှန်းကိန်းများ) တွေကို မော်ဒယ်ထဲမှာ indicator-coded အဖြစ် သတ်မှတ်ထားတဲ့အတွက်၊ factors တွေပါဝင်တဲ့ effects တွေမှာ ပုံမှန်အားဖြင့် ဆက်စပ်နေတဲ့ coefficients အများကြီး ပါဝင်နေတတ်ပါတယ်။ ပိုလျှံနေတဲ့ (redundant သို့မဟုတ် reference)

parameter နဲ့ သက်ဆိုင်တဲ့ category ကလွဲရင် ကျန်တဲ့ category တစ်ခုချင်းစီအတွက် coefficient တစ်ခုစီ ရှိနေမှာပါ။

Styles (စတိုင်များ)

Style dropdown list ကနေ ရွေးချယ်ဝင်ရောက်နိုင်တဲ့ မတူညီတဲ့ ပြသမှုပုံစံတွေ ရှိပါတယ်-

- **Diagram (ပုံကြမ်း)။** ဒါက intercept ကို ပထမဆုံး ပြသပေးပြီး၊ အဲဒီနောက်မှာ effects တွေကို predictor importance (ခန့်မှန်းကိန်း အရေးပါမှု) အကြီးမှအငယ်အတိုင်း အပေါ်ကနေအောက် စီစဉ်ပြသထားတဲ့ chart တစ်ခု ဖြစ်ပါတယ်။ Factors တွေပါဝင်တဲ့ effects တွေအတွင်းမှာတော့ coefficients တွေကို data values တွေရဲ့ အငယ်မှအကြီး (ascending order) အတိုင်း စီစဉ်ထားပါတယ်။ ပုံထဲက ဆက်သွယ်ထားတဲ့ မျဉ်း (connecting lines) တွေကို coefficient ရဲ့ လက္ခဏာ (sign) အပေါ် မူတည်ပြီး အရောင်ခြယ်ထားကာ (diagram key ကို ကြည့်ပါ)၊ coefficient ရဲ့ significance အပေါ် မူတည်ပြီး အလေးချိန်ပေး (weighted) ထားပါတယ်။ မျဉ်းအထူ (line width) ပိုကြီးလေ ပိုပြီး အရေးပါတဲ့ (more significant ဖြစ်တဲ့ - ဆိုလိုတာက p-values ပိုသေးငယ်တဲ့) coefficients တွေကို ကိုယ်စားပြုလေ ဖြစ်ပါတယ်။ ဆက်သွယ်ထားတဲ့ မျဉ်းပေါ်ကို mouse တင် (hover) လိုက်မယ်ဆိုရင် coefficient ရဲ့ တန်ဖိုး၊ သူ့ရဲ့ p-value နဲ့ အဲဒီ parameter ပါဝင်ဆက်စပ်နေတဲ့ effect ရဲ့ အရေးပါမှု (importance) တွေကို ပြသပေးတဲ့ tooltip လေး ပေါ်လာမှာပါ။ ဒါက ပုံမှန် (default) အပြင်အဆင် ဖြစ်ပါတယ်။
- **Table (ဇယား)။** ဒါက မော်ဒယ်ရဲ့ သီးခြား coefficient တစ်ခုချင်းစီအတွက် တန်ဖိုးများ၊ significance စမ်းသပ်ချက်များ နဲ့ confidence intervals တွေကို ပြသပေးပါတယ်။ Intercept ပြီးတဲ့နောက်မှာ effects တွေကို predictor importance အကြီးမှအငယ်အတိုင်း အပေါ်ကနေအောက် စီစဉ်ထားပါတယ်။ Factors တွေပါဝင်တဲ့ effects တွေအတွင်းမှာတော့ coefficients တွေကို data values တွေရဲ့ အငယ်မှအကြီးအတိုင်း စီစဉ်ထားပါတယ်။ သတိပြုရမယ့်အချက်က ပုံမှန်အားဖြင့် ဒီဇယားကို မော်ဒယ် parameter တစ်ခုချင်းစီရဲ့ coefficient, significance နဲ့ importance တွေကိုပဲ ပြသဖို့ ခေါက်ထား (collapsed) ပေးထားပါတယ်။ Standard error, t statistic နဲ့ confidence interval တွေကို ကြည့်ချင်ရင်တော့ ဇယားထဲက Coefficient ဆဲလ် (cell) ကို နှိပ်ပါ။ ဇယားထဲက မော်ဒယ် parameter တစ်ခုရဲ့ နာမည်ပေါ်ကို mouse တင်လိုက်ရင်၊ parameter ရဲ့ နာမည်၊ အဲဒီ parameter ပါဝင်ဆက်စပ်နေတဲ့ effect၊ နဲ့ (categorical predictors တွေအတွက်ဆိုရင်) အဲဒီ မော်ဒယ် parameter နဲ့ ဆက်စပ်နေတဲ့ value labels တွေကို ပြသပေးတဲ့ tooltip လေး ပေါ်လာပါလိမ့်မယ်။ Automatic data preparation (အလိုအလျောက် data ပြင်ဆင်ခြင်း) က categorical predictor တစ်ခုရဲ့ ဆင်တူတဲ့ categories တွေကို ပေါင်းစည်း (merge) လိုက်တဲ့အခါ အသစ်ဖန်တီးလိုက်တဲ့ categories တွေကို ကြည့်ရှုဖို့အတွက် ဒါဟာ အထူးသဖြင့် အသုံးဝင်ပါတယ်။

Predictor importance (ခန့်မှန်းကိန်း အရေးပါမှု)

မြင်ကွင်းမှာ ဘယ် predictors တွေကို ပြသမလဲဆိုတာကို ထိန်းချုပ်ပေးတဲ့ Predictor Importance ဆလိုက်ဒါ (slider) တစ်ခု ရှိပါတယ်။ ဒီအရာက မော်ဒယ်ကို ပြောင်းလဲသွားစေတာ မဟုတ်ဘဲ၊ အရေးအပါဆုံး predictors တွေအပေါ်မှာပဲ သင့်ကို အာရုံစိုက်နိုင်စေဖို့ ကူညီပေးတာပါ။ ပုံမှန်အားဖြင့်တော့ ထိပ်ဆုံးက အရေးအပါဆုံး effects ၁၀ ခု (top 10 effects) ကို ပြသပေးထားပါတယ်။

Significance (အရေးပါမှု)

Predictor importance အပေါ်အခြေခံပြီး ပြသထားတာတွေအပြင်၊ ဘယ် coefficients တွေကို မြင်ကွင်းမှာ ဆက်လက်ပြသမလဲဆိုတာကို ထပ်မံထိန်းချုပ်ပေးတဲ့ Significance ဆလိုက်ဒါ (slider) တစ်ခု ရှိပါတယ်။ ဆလိုက်ဒါမှာ သတ်မှတ်ထားတဲ့တန်ဖိုးထက် ကြီးတဲ့ significance values ရှိတဲ့ coefficients တွေကို ဖျောက်ထား (hidden) ပေးပါလိမ့်မယ်။ ဒါကလည်း မော်ဒယ်ကို ပြောင်းလဲသွားစေတာ မဟုတ်ဘဲ၊ အရေးအပါဆုံး coefficients တွေအပေါ်မှာပဲ သင့်ကို အာရုံစိုက်နိုင်စေဖို့ ကူညီပေးတာပါ။ ပုံမှန်အားဖြင့် default တန်ဖိုးက 1.00 ဖြစ်တဲ့အတွက် significance အပေါ် မူတည်ပြီး ဘယ် coefficients ကိုမှ စစ်ထုတ် (filtered) ဖျောက်ထားမှာ မဟုတ်ပါဘူး။

Estimated Means (ခန့်မှန်း ပျမ်းမျှကိန်းများ)

ဒါတွေက significant ဖြစ်တဲ့ (အရေးပါတဲ့) predictors တွေအတွက် ပြသပေးတဲ့ charts တွေ ဖြစ်ပါတယ်။ အခြား predictors အားလုံးကို ကိန်းသေ (constant) အဖြစ် ထားရှိပြီး၊ horizontal axis (အလျားလိုက်ဝင်ရိုး) မှာရှိတဲ့ predictor ရဲ့ တန်ဖိုးတစ်ခုချင်းစီအတွက် vertical axis (ဒေါင်လိုက်ဝင်ရိုး) ပေါ်က target ရဲ့ မော်ဒယ်-ခန့်မှန်းတန်ဖိုး (model-estimated value) ကို ဒီ chart က ပြသပေးပါတယ်။ ၎င်းက predictor တစ်ခုချင်းစီရဲ့ coefficients တွေဟာ target အပေါ် ဘယ်လိုသက်ရောက်မှုရှိသလဲဆိုတာကို အသုံးဝင်တဲ့ အမြင်အာရုံပုံဖော်မှု (visualization) တစ်ခုအနေနဲ့ ထောက်ပံ့ပေးပါတယ်။

မှတ်ချက်- တကယ်လို့ ဘယ် predictors မှ significant မဖြစ်ဘူးဆိုရင်တော့ estimated means တွေကို ထုတ်ပေးမှာ မဟုတ်ပါဘူး။

Model Building Summary (မော်ဒယ်တည်ဆောက်မှု အကျဉ်းချုပ်)

Model Selection အပြင်အဆင်များမှာ "None" (ဘာမှမရွေးချယ်ခြင်း) ကလွဲပြီး အခြား မော်ဒယ်ရွေးချယ်တဲ့ algorithm တစ်ခုခုကို ရွေးချယ်ထားမယ်ဆိုရင်၊ ဒီမြင်ကွင်းက မော်ဒယ်တည်ဆောက်မှု လုပ်ငန်းစဉ်ရဲ့ အသေးစိတ်အချက်အလက် တချို့ကို ဖော်ပြပေးပါတယ်။

- **Forward stepwise (ရှေ့သို့ တစ်ဆင့်ချင်း ရွေးချယ်ခြင်း)။** Forward stepwise ကို ရွေးချယ်တဲ့ algorithm အဖြစ် သတ်မှတ်ထားတဲ့အခါ၊ ဇယားက stepwise algorithm ရဲ့ နောက်ဆုံး အဆင့် ၁၀ ဆင့် (last 10 steps) ကို ပြသပေးပါတယ်။ အဆင့်တစ်ဆင့်ချင်းစီအတွက်၊ ရွေးချယ်မှု စံသတ်မှတ်ချက် (selection criterion) ရဲ့ တန်ဖိုးနဲ့ အဲဒီအဆင့်မှာ မော်ဒယ်ထဲ ပါဝင်နေတဲ့ သက်ရောက်မှု (effects) တွေကို ပြသပေးပါတယ်။ ဒါက အဆင့်တစ်ဆင့်ချင်းစီဟာ မော်ဒယ်အတွက် ဘယ်လောက်အထိ အထောက်အကူပြု ပုံပိုးပေးသလဲဆိုတာကို သင့်ကို နားလည်သဘောပေါက်စေပါတယ်။ ပေးထားတဲ့ အဆင့်တစ်ခုမှာ ဘယ် effects တွေက မော်ဒယ်ထဲမှာ ပါဝင်နေသလဲဆိုတာကို ပိုပြီး လွယ်ကူစွာ မြင်တွေ့နိုင်ဖို့အတွက်၊ column တစ်ခုချင်းစီကနေ rows တွေကို အစဉ်လိုက်စီ (sort) လို့ရအောင် ပြုလုပ်ပေးထားပါတယ်။

- **Best subsets (အကောင်းဆုံး အုပ်စုငယ်များ ရွေးချယ်ခြင်း)။** Best subsets ကို ရွေးချယ်တဲ့ algorithm အဖြစ် သတ်မှတ်ထားတဲ့အခါ၊ ဇယားက ထိပ်ဆုံး (အကောင်းဆုံး) မော်ဒယ် ၁၀ ခုကို ပြသပေးပါတယ်။ မော်ဒယ်တစ်ခုချင်းစီအတွက်၊ ရွေးချယ်မှု စံသတ်မှတ်ချက်ရဲ့ တန်ဖိုးနဲ့ မော်ဒယ်ထဲမှာ ပါဝင်တဲ့ effects တွေကို ပြသပေးပါတယ်။ ဒါက ထိပ်ဆုံး မော်ဒယ်တွေရဲ့ တည်ငြိမ်မှု (stability) ကို သင့်ကို နားလည်သဘောပေါက်စေပါတယ်။ တကယ်လို့ အဲဒီမော်ဒယ်တွေကြားမှာ ကွာခြားချက် နည်းနည်းပဲရှိပြီး ဆင်တူတဲ့ effects တွေ အများကြီး ပါဝင်နေမယ်ဆိုရင်၊ သင်က အဲဒီ "ထိပ်ဆုံး" ("top") မော်ဒယ်အပေါ်မှာ အတော်လေး ယုံကြည်စိတ်ချနိုင်ပါတယ်။ တကယ်လို့ သူတို့ဆီမှာ အရမ်းကို ကွဲပြားခြားနားတဲ့ effects တွေ ပါဝင်နေမယ်ဆိုရင်တော့၊ effects တချို့က အရမ်းကို ဆင်တူနေနိုင်တဲ့အတွက် သူတို့ကို ပေါင်းစည်းပစ်သင့်ပါတယ် (ဒါမှမဟုတ် တစ်ခုကို ဖယ်ရှားပစ်သင့်ပါတယ်)။ ပေးထားတဲ့ အဆင့်တစ်ခုမှာ ဘယ် effects တွေက မော်ဒယ်ထဲမှာ ပါဝင်နေသလဲဆိုတာကို ပိုပြီး လွယ်ကူစွာ မြင်တွေ့နိုင်ဖို့အတွက်၊ column တစ်ခုချင်းစီကနေ rows တွေကို အစဉ်လိုက်စီ (sort) လို့ရအောင် ပြုလုပ်ပေးထားပါတယ်။

Linear Regression (မျဉ်းဖြောင့် ဆက်စပ်မှု ခွဲခြမ်းစိတ်ဖြာခြင်း)

Linear Regression ဟာ မှီခိုပြောင်းလဲကိန်း (dependent variable) ရဲ့ တန်ဖိုးကို အကောင်းဆုံး ခန့်မှန်းပေးနိုင်ဖို့အတွက် လွတ်လပ်သော ပြောင်းလဲကိန်း (independent variables) တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ပါဝင်တဲ့ မျဉ်းဖြောင့်ညီမျှခြင်း (linear equation) ရဲ့ မြှောက်ဖော်ကိန်း (coefficients) တွေကို ခန့်မှန်းတွက်ချက်ပေးပါတယ်။ ဥပမာအနေနဲ့ သင်က အရောင်းကိုယ်စားလှယ်တစ်ယောက်ရဲ့ တစ်နှစ်စာ စုစုပေါင်း အရောင်းပမာဏ (dependent variable) ကို သူ့ရဲ့ အသက်၊ ပညာအရည်အချင်း နဲ့ လုပ်သက် အစရှိတဲ့ independent variables တွေအပေါ် အခြေခံပြီး ခန့်မှန်းကြည့်လို့ ရပါတယ်။

Example (ဥပမာ)

ဘတ်စကက်ဘောအသင်းတစ်သင်း တစ်ရာသီမှာ နိုင်တဲ့ပွဲအရေအတွက်ဟာ အဲဒီအသင်း တစ်ပွဲမှာ ပျမ်းမျှရရှိတဲ့ အမှတ်အရေအတွက်နဲ့ ဆက်စပ်မှု ရှိပါသလား။ Scatterplot (အစက်ချဇယား) အရ ဒီပြောင်းလဲကိန်း နှစ်ခုဟာ မျဉ်းဖြောင့်ဆက်စပ်မှု (linear relationship) ရှိတယ်လို့ ညွှန်ပြနေပါတယ်။ နိုင်ပွဲအရေအတွက်နဲ့ ပြိုင်ဘက်အသင်းက ရရှိတဲ့ ပျမ်းမျှအမှတ်အရေအတွက် တို့ဟာလည်း မျဉ်းဖြောင့်ဆက်စပ်မှု ရှိကြပါတယ်။ အဲဒီ variables နှစ်ခုကတော့ ပြောင်းပြန်ဆက်စပ်မှု (negative relationship) ရှိကြတာပါ။ နိုင်ပွဲအရေအတွက် တိုးလာလေလေ၊ ပြိုင်ဘက်က ရတဲ့ ပျမ်းမျှအမှတ်အရေအတွက်က လျော့နည်းသွားလေလေ ဖြစ်ပါတယ်။ Linear regression ကို အသုံးပြုပြီး ဒီ variables တွေရဲ့ ဆက်စပ်မှုကို မော်ဒယ် တည်ဆောက်နိုင်ပါတယ်။ ကောင်းမွန်တဲ့ မော်ဒယ်တစ်ခုကို အသုံးပြုပြီး အသင်းတွေအနေနဲ့ ပွဲအရေအတွက် ဘယ်လောက် နိုင်မလဲဆိုတာကို ခန့်မှန်းနိုင်ပါတယ်။

Statistics (စာရင်းအင်းဆိုင်ရာ အချက်အလက်များ)

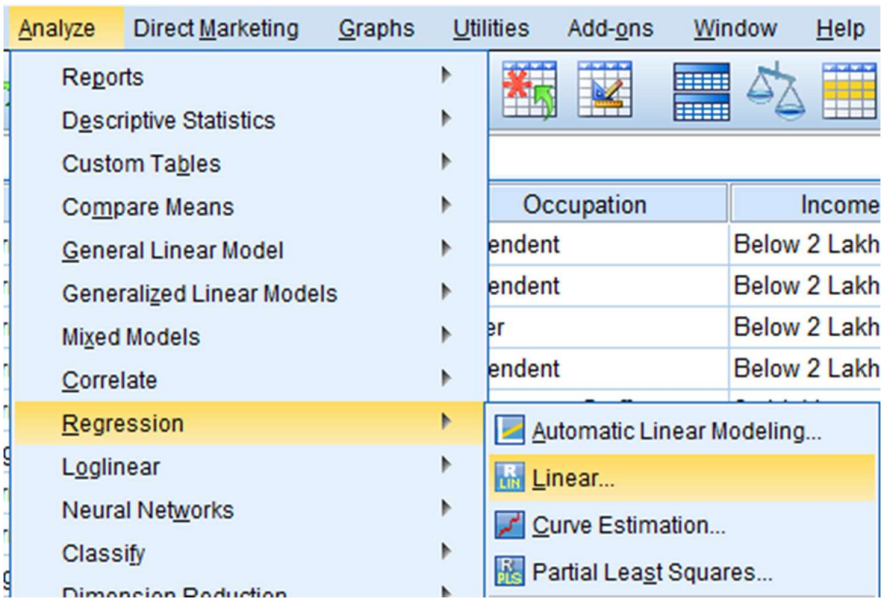
Variable တစ်ခုချင်းစီအတွက်- မှန်ကန်တဲ့ cases အရေအတွက်၊ mean (ပျမ်းမျှကိန်း) နဲ့ standard deviation (စံသွေဖည်ကိန်း) တို့ ပါဝင်ပါတယ်။ မော်ဒယ်တစ်ခုချင်းစီအတွက်- regression coefficients, correlation matrix, part and partial correlations, multiple R, R^2, adjusted R^2, change in R^2, standard error of the estimate, analysis-of-variance table, predicted values, နဲ့ residuals တို့ ပါဝင်ပါတယ်။ ဒါ့အပြင် regression coefficient တစ်ခုချင်းစီအတွက် 95%-confidence intervals၊ variance-covariance matrix, variance inflation factor, tolerance, Durbin-Watson test, distance measures (Mahalanobis, Cook, နဲ့ leverage values)၊ DfBeta, DfFit, prediction intervals နဲ့ casewise diagnostic information တွေကိုလည်း ရရှိနိုင်ပါတယ်။ ဇယား (Plots) တွေအနေနဲ့- scatterplots, partial plots, histograms နဲ့ normal probability plots တွေကို ရရှိနိုင်ပါတယ်။

Linear Regression Data Considerations (Linear Regression အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ)

- **Data (အချက်အလက်)။** Dependent နဲ့ independent variables တွေဟာ အရေအတွက်ပြ (quantitative) ဖြစ်သင့်ပါတယ်။ ကိုးကွယ်ရာဘာသာ၊ သင်ယူခဲ့တဲ့ ဘာသာရပ်၊ ဒါမှမဟုတ် နေထိုင်ရာဒေသ လိုမျိုး အမျိုးအစားခွဲ (categorical) variables တွေကိုတော့ binary (dummy) variables တွေ ဒါမှမဟုတ် အခြား contrast variables ပုံစံမျိုးတွေအဖြစ် ပြန်လည်ပြောင်းလဲ (recode) ဖို့ လိုအပ်ပါတယ်။
- **Assumptions (ယူဆချက်များ)။** Independent variable ရဲ့ တန်ဖိုးတစ်ခုချင်းစီအတွက်၊ dependent variable ရဲ့ ဖြန့်ကျက်မှု (distribution) ဟာ normal distribution ဖြစ်ရပါမယ်။ Dependent variable ဖြန့်ကျက်မှုရဲ့ variance ဟာ independent variable ရဲ့ တန်ဖိုးအားလုံးအတွက် ကိန်းသေ (constant) ဖြစ်သင့်ပါတယ်။ Dependent variable နဲ့ independent variable တစ်ခုချင်းစီကြားက ဆက်စပ်မှုဟာ မျဉ်းဖြောင့် (linear) ဖြစ်သင့်ပြီး၊ လေ့လာတွေ့ရှိချက် (observations) အားလုံးဟာ တစ်ခုနဲ့တစ်ခု အမှီအခိုကင်း (independent) ဖြစ်ရပါမယ်။

To Obtain a Linear Regression Analysis (Linear Regression ခွဲခြမ်းစိတ်ဖြာမှုကို ရယူအသုံးပြုရန်)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Regression > Linear...



၂။ Linear Regression dialog box ထဲမှာ ဂဏန်း (numeric) dependent variable တစ်ခုကို ရွေးချယ်ပါ။

၃။ ဂဏန်း (numeric) independent variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။

ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့ သင် အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- Independent variables တွေကို အုပ်စု (blocks) တွေအဖြစ် ခွဲခြားနိုင်ပြီး၊ variables အုပ်စုကွဲတွေအတွက် မတူညီတဲ့ entry methods တွေကို သတ်မှတ်နိုင်ပါတယ်။
- သီးခြားတန်ဖိုးတစ်ခု (ဒါမှမဟုတ် တန်ဖိုးများ) ရှိတဲ့ cases တွေအတွက်သာ ခွဲခြမ်းစိတ်ဖြာမှုကို ကန့်သတ်ဖို့ selection variable တစ်ခုကို ရွေးချယ်နိုင်ပါတယ်။
- ဇယား (plots) တွေပေါ်က အမှတ် (points) တွေကို ခွဲခြားသိရှိနိုင်ဖို့ case identification variable တစ်ခုကို ရွေးချယ်နိုင်ပါတယ်။
- Weighted least squares analysis အတွက် ဂဏန်း (numeric) WLS Weight variable တစ်ခုကို ရွေးချယ်နိုင်ပါတယ်။

WLS (Weighted Least Squares) သင့်ကို weighted least-squares model တစ်ခု ရရှိစေဖို့ ခွင့်ပြုပါတယ်။ Data points တွေကို သူတို့ရဲ့ variances တွေရဲ့ ပြောင်းပြန်ကိန်း (reciprocal) နဲ့ အလေးချိန် (weight) ပေးထားပါတယ်။ ဆိုလိုတာက ကြီးမားတဲ့ variances ရှိတဲ့ observations တွေဟာ သေးငယ်တဲ့ variances ရှိတဲ့ observations တွေထက်စာရင် analysis အပေါ် သက်ရောက်မှု ပိုနည်းပါတယ်။ တကယ်လို့ weighting variable ရဲ့ တန်ဖိုးက သုည၊ အနှုတ်လက္ခဏာ ဒါမှမဟုတ် missing ဖြစ်နေမယ်ဆိုရင်၊ အဲဒီ case ကို analysis ထဲကနေ ဖယ်ထုတ်ထားမှာ ဖြစ်ပါတယ်။

Linear Regression Variable Selection Methods (Linear Regression ပြောင်းလဲကိန်းများ ရွေးချယ်သည့် နည်းလမ်းများ)

Method (နည်းလမ်း) ရွေးချယ်ခြင်းက independent variables တွေကို analysis (ခွဲခြမ်းစိတ်ဖြာမှု) ထဲ ဘယ်လို ထည့်သွင်းမလဲဆိုတာကို သတ်မှတ်ပေးနိုင်ပါတယ်။ မတူညီတဲ့ နည်းလမ်းတွေကို အသုံးပြုခြင်းအားဖြင့်၊ တူညီတဲ့ variable အစုအဝေးတစ်ခုတည်းကနေပဲ အမျိုးမျိုးကွဲပြားတဲ့ regression models တွေကို တည်ဆောက်နိုင်ပါတယ်။

- **Enter (Regression)။** Block တစ်ခုထဲမှာရှိတဲ့ variables တွေ အားလုံးကို အဆင့်တစ်ဆင့်တည်း (single step) နဲ့ တစ်ပြိုင်နက်တည်း ထည့်သွင်းပေးတဲ့ variable ရွေးချယ်မှု နည်းလမ်းတစ်ခု ဖြစ်ပါတယ်။
- **Stepwise။** အဆင့်တစ်ဆင့်ချင်းစီတိုင်းမှာ၊ ညီမျှခြင်း (equation) ထဲမှာ မပါဝင်သေးတဲ့ independent variable တွေထဲကမှ F ရဲ့ probability အငယ်ဆုံးဖြစ်နေတဲ့ variable ကို

(အဲဒီ probability က လုံလောက်အောင် ငယ်နေမယ်ဆိုရင်) ထည့်သွင်းပေးပါတယ်။ Regression equation ထဲမှာ ပါဝင်ပြီးသား variables တွေရဲ့ F probability က အရမ်းကြီးသွားမယ်ဆိုရင်တော့ သူတို့ကို ပြန်လည်ဖယ်ရှား (remove) ပစ်ပါတယ်။ ထည့်သွင်းဖို့ ဒါမှမဟုတ် ဖယ်ရှားဖို့ ကိုက်ညီတဲ့ variables တွေ မရှိတော့တဲ့အခါမှာတော့ ဒီလုပ်ငန်းစဉ် ရပ်တန့်သွားမှာ ဖြစ်ပါတယ်။

- **Remove။** Block တစ်ခုထဲမှာရှိတဲ့ variables တွေ အားလုံးကို အဆင့်တစ်ဆင့်တည်းနဲ့ တစ်ပြိုင်နက်တည်း ဖယ်ရှားပစ်တဲ့ variable ရွေးချယ်မှု နည်းလမ်းတစ်ခု ဖြစ်ပါတယ်။

- **Backward Elimination။** Variables တွေ အားလုံးကို equation ထဲ အရင်ဆုံး ထည့်သွင်းလိုက်ပြီးမှ တစ်ခုချင်းစီ အစဉ်လိုက် ပြန်လည်ဖယ်ရှားတဲ့ နည်းလမ်းဖြစ်ပါတယ်။ Dependent variable နဲ့ partial correlation အငယ်ဆုံးဖြစ်နေတဲ့ variable ကို ဖယ်ရှားဖို့အတွက် ပထမဆုံး စဉ်းစားပါတယ်။ တကယ်လို့ သူက ဖယ်ရှားဖို့ သတ်မှတ်ချက် (criterion) နဲ့ ကိုက်ညီနေရင် ဖယ်ရှားပစ်ပါတယ်။ ပထမဆုံး variable ကို ဖယ်ရှားပြီးတဲ့နောက်မှာ၊ equation ထဲမှာ ကျန်နေတဲ့ variables တွေထဲကမှ partial correlation အငယ်ဆုံးဖြစ်တဲ့ variable ကို ဆက်လက်စဉ်းစားပါတယ်။ ဖယ်ရှားဖို့ သတ်မှတ်ချက်တွေနဲ့ ကိုက်ညီတဲ့ variables တွေ equation ထဲမှာ မရှိတော့တဲ့အခါမှာ လုပ်ငန်းစဉ် ရပ်တန့်သွားမှာ ဖြစ်ပါတယ်။

- **Forward Selection။** ဒါကတော့ variables တွေကို မော်ဒယ်ထဲ အစဉ်လိုက် တစ်ခုချင်းစီ ထည့်သွင်းသွားတဲ့ stepwise variable ရွေးချယ်မှု လုပ်ငန်းစဉ်တစ်ခုပါ။ Equation ထဲ ထည့်သွင်းဖို့ ပထမဆုံး စဉ်းစားခံရမယ့် variable ကတော့ dependent variable နဲ့ အကြီးမားဆုံး အပေါင်း သို့မဟုတ် အနှုတ် correlation (largest positive or negative correlation) ရှိတဲ့ variable ပဲ ဖြစ်ပါတယ်။ အဲဒီ variable ဟာ ထည့်သွင်းဖို့ သတ်မှတ်ချက် (criterion) နဲ့ ကိုက်ညီမှသာ equation ထဲကို ထည့်သွင်းမှာပါ။ ပထမဆုံး variable ကို ထည့်သွင်းပြီးသွားရင်တော့၊ equation ထဲမှာ မပါဝင်သေးတဲ့ independent variables တွေထဲကမှ partial correlation အကြီးဆုံးဖြစ်နေတဲ့ ကောင်ကို ဆက်လက်စဉ်းစားပါတယ်။ ထည့်သွင်းဖို့ သတ်မှတ်ချက်နဲ့ ကိုက်ညီတဲ့ variables တွေ မရှိတော့တဲ့အခါ လုပ်ငန်းစဉ် ရပ်တန့်သွားပါလိမ့်မယ်။

သင်ရရှိလာတဲ့ output ထဲက significance values (အရေးပါမှုတန်ဖိုးများ) ဟာ မော်ဒယ်တစ်ခုတည်း (single model) ကို fitting လုပ်ထားခြင်းအပေါ် အခြေခံထားတာပါ။ ဒါကြောင့်မို့လို့ stepwise, forward, ဒါမှမဟုတ် backward လိုမျိုး အဆင့်လိုက်ရွေးချယ်တဲ့ (stepwise) နည်းလမ်းတစ်ခုခုကို သုံးထားမယ်ဆိုရင် အဲဒီ significance values တွေက ယေဘုယျအားဖြင့် အကျိုးမဝင် (invalid) ပါဘူး။

ဘယ်လို ထည့်သွင်းမယ့်နည်းလမ်း (entry method) ကိုပဲ သတ်မှတ်ထားပါစေ၊ variables တွေအားလုံးဟာ equation ထဲ ဝင်ရောက်နိုင်ဖို့အတွက် tolerance criterion (ခံနိုင်ရည်ဆိုင်ရာ သတ်မှတ်ချက်) ကို အောင်မြင်စွာ ဖြတ်ကျော်နိုင်ရပါမယ်။ ပုံမှန် (default) tolerance level ကတော့ 0.0001 ဖြစ်ပါတယ်။ ဒါ့အပြင်၊ variable တစ်ခုကို ထည့်သွင်းလိုက်ခြင်းအားဖြင့် မော်ဒယ်ထဲမှာ ပါဝင်ပြီးသား တခြား variable တစ်ခုရဲ့ tolerance ကို သတ်မှတ်ထားတဲ့ tolerance criterion အောက် ကျဆင်းသွားစေမယ်ဆိုရင် အဲဒီ variable ကို ထည့်သွင်းမှာ မဟုတ်ပါဘူး။

ရွေးချယ်ထားတဲ့ independent variables တွေ အားလုံးကို regression model တစ်ခုတည်းထဲကိုပဲ ပေါင်းထည့်ပေးတာ ဖြစ်ပါတယ်။ ဒါပေမယ့် variables အုပ်စုကွဲလေးတွေ (subsets) အလိုက် မတူညီတဲ့ ထည့်သွင်းမှု နည်းလမ်း (entry methods) တွေကို သင်ကိုယ်တိုင် သတ်မှတ်ပေးလို့ ရပါတယ်။ ဥပမာအားဖြင့်၊ သင်က regression model ထဲကို variable တစ်အုပ်စု (one block of variables) ကို stepwise selection သုံးပြီး ထည့်သွင်းကာ၊ ဒုတိယအုပ်စု (second block) ကိုတော့ forward selection သုံးပြီး ထည့်သွင်းတာမျိုး လုပ်လို့ရပါတယ်။ Regression model ထဲကို ဒုတိယမြောက် variable block တစ်ခု ထပ်ထည့်ချင်ရင်တော့ **Next** ကို နှိပ်နိုင်ပါတယ်။

Linear Regression Set Rule (Linear Regression ပြောင်းလဲကိန်း သတ်မှတ်ချက် စည်းမျဉ်း)

Selection rule က သတ်မှတ်ပေးထားတဲ့ အခြေအနေနဲ့ ကိုက်ညီတဲ့ cases တွေကိုပဲ analysis မှာ ထည့်သွင်းတွက်ချက်ပေးမှာပါ။ ဥပမာ- သင်က variable တစ်ခုကို ရွေးပြီး၊ equals (ညီမျှသည်) ကို ရွေးကာ၊ တန်ဖိုးနေရာမှာ 5 လို့ ရိုက်ထည့်လိုက်မယ်ဆိုရင်၊ အဲဒီရွေးချယ်ထားတဲ့ variable ရဲ့ တန်ဖိုး 5 နဲ့ ညီမျှတဲ့ cases တွေကိုပဲ analysis ထဲ ထည့်သွင်းတွက်ချက်ပေးမှာ ဖြစ်ပါတယ်။ ဒီနေရာမှာ ဂဏန်းတွေသာမက စာသား (string) တန်ဖိုးတွေကိုလည်း အသုံးပြုခွင့် ရှိပါတယ်။

Linear Regression Plots (Linear Regression ပုံပြဇယားများ)

Plots (ဇယား) တွေက normality, linearity နဲ့ equality of variances စတဲ့ ယူဆချက် (assumptions) တွေ မှန်ကန်မှုရှိမရှိ စစ်ဆေးအတည်ပြုရာမှာ အကူအညီပေးပါတယ်။ Outliers တွေ၊ ပုံမှန်မဟုတ်တဲ့ လေ့လာတွေ့ရှိချက်တွေနဲ့ မော်ဒယ်အပေါ် လွှမ်းမိုးမှုရှိတဲ့ cases တွေကို ရှာဖွေရာမှာလည်း plots တွေက အရမ်းအသုံးဝင်ပါတယ်။ ၎င်းတို့ကို variables အသစ်တွေအနေနဲ့ save လုပ်ပြီးတဲ့နောက်မှာတော့၊ independent variables တွေနဲ့ plots တွေဆွဲဖို့အတွက် Data Editor ထဲမှာ predicted values, residuals နဲ့ တခြား diagnostic အချက်အလက်တွေကို ရရှိနိုင်ပါပြီ။ အောက်ပါ plots တွေကို ရရှိနိုင်ပါတယ်-

- **Scatterplots (အစက်ချဇယားများ)။** သင်က အောက်ပါတို့ထဲက ကြိုက်ရာနှစ်ခုကို တွဲပြီး plot ဆွဲနိုင်ပါတယ်- dependent variable, standardized predicted values,

standardized residuals, deleted residuals, adjusted predicted values, Studentized residuals သို့မဟုတ် Studentized deleted residuals။ Linearity နဲ့ equality of variances ကို စစ်ဆေးဖို့ဆိုရင်တော့ standardized residuals နဲ့ standardized predicted values တို့ကို ယှဉ်ပြီး plot ဆွဲကြည့်ပါ။

- **Source variable list (မူရင်း ပြောင်းလဲကိန်း စာရင်း)** ။ ၎င်းက dependent variable (DEPENDNT) အပြင် အောက်ပါ predicted နဲ့ residual variables တွေကို စာရင်းပြပေးပါတယ်- Standardized predicted values (*ZPRED), Standardized residuals (*ZRESID), Deleted residuals (*DRESID), Adjusted predicted values (*ADJPRED), Studentized residuals (*SRESID) နဲ့ Studentized deleted residuals (*SDRESID) တို့ ဖြစ်ပါတယ်။
- **Produce all partial plots (တစ်စိတ်တစ်ပိုင်း ဇယားများအားလုံး ထုတ်ရန်)** ။ Independent variable တစ်ခုချင်းစီရဲ့ residuals တွေနဲ့ dependent variable ရဲ့ residuals တွေကို (ကျန်တဲ့ independent variables တွေအပေါ် သီးခြားစီ regress လုပ်ပြီးတဲ့နောက်) scatterplots အနေနဲ့ ပြသပေးပါတယ်။ Partial plot တစ်ခု ထုတ်ဖို့ဆိုရင် ညီမျှခြင်းထဲမှာ independent variables အနည်းဆုံး နှစ်ခု ပါဝင်နေရပါမယ်။
- **Standardized Residual Plots**။ Standardized residuals တွေရဲ့ ဟစ္စတိုဂရမ် (histograms) တွေနဲ့၊ standardized residuals တွေရဲ့ ဖြန့်ကျက်မှုကို normal distribution နဲ့ နှိုင်းယှဉ်ပြတဲ့ normal probability plots တွေကို ရယူနိုင်ပါတယ်။

တကယ်လို့ သင်က plots တွေတောင်းဆိုထားခဲ့မယ်ဆိုရင်၊ standardized predicted values နဲ့ standardized residuals (*ZPRED နဲ့ *ZRESID) တို့အတွက် summary statistics တွေကိုပါ ပြသပေးမှာ ဖြစ်ပါတယ်။

Linear Regression: Saving New Variables (ပြောင်းလဲကိန်းအသစ်များ သိမ်းဆည်းခြင်း)

Predicted values, residuals နဲ့ ရောဂါရှာဖွေရေး အချက်အလက် (diagnostic information) အဖြစ် အသုံးဝင်တဲ့ အခြား statistics တွေကို save လုပ်နိုင်ပါတယ်။ ရွေးချယ်မှုတစ်ခုချင်းစီက သင့်ရဲ့ active data file ထဲကို variable အသစ် တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ပေါင်းထည့်ပေးသွားမှာပါ။

- **Predicted Values (ခန့်မှန်းတန်ဖိုးများ)**။ Regression model က case တစ်ခုချင်းစီအတွက် ခန့်မှန်းပေးတဲ့ တန်ဖိုးတွေ ဖြစ်ပါတယ်။
    - ***Unstandardized.*** Dependent variable အတွက် မော်ဒယ်က ခန့်မှန်းပေးတဲ့ မူလတန်ဖိုး။

- ***Standardized.*** Predicted value တစ်ခုချင်းစီကို သူ့ရဲ့ စံသတ်မှတ်ချက်ပုံစံ (standardized form) အဖြစ် ပြောင်းလဲထားတာပါ။
- ***Adjusted.*** Regression coefficients တွေ တွက်ချက်မှုထဲကနေ အဲဒီ case ကို ဖယ်ထုတ်လိုက်တဲ့အခါ ထွက်လာမယ့် predicted value ပါ။
- ***S.E. of mean predictions.*** Predicted values တွေရဲ့ Standard errors ပါ။

- **Distances (အကွာအဝေးများ)။** Independent variables တွေအတွက် ပုံမှန်မဟုတ်တဲ့ တန်ဖိုးတွဲတွေရှိတဲ့ cases တွေနဲ့ regression မော်ဒယ်အပေါ် ကြီးမားတဲ့ သက်ရောက်မှုရှိနိုင်တဲ့ cases တွေကို ခွဲခြားသိရှိနိုင်ဖို့ တိုင်းတာမှုတွေ ဖြစ်ပါတယ်။
  - ***Mahalanobis.*** Case တစ်ခုရဲ့ independent variables တန်ဖိုးတွေက cases အားလုံးရဲ့ ပျမ်းမျှနဲ့ ဘယ်လောက်ကွာခြားသလဲဆိုတာကို တိုင်းတာတာပါ။
  - ***Cook's.*** Case တစ်ခုကို ဖယ်ထုတ်လိုက်ရင် regression coefficients တွေ ဘယ်လောက်ပြောင်းလဲသွားမလဲ တိုင်းတာတာပါ။
  - ***Leverage values.*** Regression ရဲ့ fit အပေါ် အမှတ်တစ်ခုစီရဲ့ လွှမ်းမိုးမှုကို တိုင်းတာပါတယ်။
- **Prediction Intervals (ခန့်မှန်းမှု ကြားကာလများ) ။** Mean နဲ့ individual prediction intervals နှစ်ခုစလုံးအတွက် အထက်နဲ့ အောက် ကန့်သတ်ချက်များ ဖြစ်ပါတယ်။
  - ***Mean.*** Mean predicted response အတွက် prediction interval ရဲ့ အထက်နဲ့ အောက် ကန့်သတ်ချက်များ။
  - ***Individual.*** Case တစ်ခုတည်းအတွက် dependent variable ရဲ့ prediction interval ရဲ့ အထက်နဲ့ အောက် ကန့်သတ်ချက်များ။
  - ***Confidence Interval.*** 1 နဲ့ 99.99 ကြား တန်ဖိုးတစ်ခုထည့်ပြီး Prediction Intervals နှစ်ခုအတွက် ယုံကြည်မှုအဆင့်ကို သတ်မှတ်ပေးနိုင်ပါတယ်။
- **Residuals (အကြွင်းအကျန်များ / ကွာဟချက်များ) ။** တွေ့ရှိရတဲ့ dependent variable တန်ဖိုးထဲကနေ regression equation အရ ခန့်မှန်းထားတဲ့ တန်ဖိုးကို နှုတ်ထားတဲ့ ရလဒ်တွေပါ။
  - ***Unstandardized.*** လေ့လာတွေ့ရှိရတဲ့ တန်ဖိုးနဲ့ မော်ဒယ်က ခန့်မှန်းပေးတဲ့ တန်ဖိုးတို့ ကွာခြားချက်ပါ။
  - ***Standardized.*** Residual ကို သူ့ရဲ့ standard deviation နဲ့ စားထားတာပါ။

- ***Studentized.*** Residual ကို case တစ်ခုနဲ့တစ်ခု မတူညီဘဲ ပြောင်းလဲနေတဲ့ standard deviation နဲ့ စားထားတာပါ။
- ***Deleted.*** Case တစ်ခုကို ဖယ်ထုတ်ပြီး တွက်ထားတဲ့ residual ပါ။
- ***Studentized deleted.*** Deleted residual ကို သူ့ရဲ့ standard error နဲ့ စားထားတာပါ။

- **Influence Statistics (လွှမ်းမိုးမှုဆိုင်ရာ စာရင်းအင်းများ)** ။ သီးခြား case တစ်ခုကို ဖယ်ထုတ်လိုက်ခြင်းကြောင့် regression coefficients တွေနဲ့ predicted values တွေအပေါ် ဖြစ်ပေါ်လာတဲ့ အပြောင်းအလဲတွေ ဖြစ်ပါတယ်။
  - ***DfBetas.*** သီးခြား case တစ်ခုကို ဖယ်ထုတ်လိုက်ခြင်းကြောင့် regression coefficient အပေါ် ဖြစ်ပေါ်လာတဲ့ ပြောင်းလဲမှုပါ။
  - ***Standardized DfBeta.*** DfBetas ကို စံသတ်မှတ်ထား (standardized) တာပါ။
  - ***DfFit.*** သီးခြား case တစ်ခုကို ဖယ်ထုတ်လိုက်ခြင်းကြောင့် predicted value အပေါ် ဖြစ်ပေါ်လာတဲ့ ပြောင်းလဲမှုပါ။
  - ***Standardized DfFit.*** DfFit ကို စံသတ်မှတ်ထားတာပါ။
  - ***Covariance ratio.*** သီးခြား case ကို ဖယ်ထုတ်တွက်ထားတဲ့ covariance matrix ရဲ့ determinant နဲ့ အားလုံးပါတဲ့ determinant တို့ရဲ့ အချိုးပါ။
- **Coefficient Statistics.** Regression coefficients တွေကို dataset ဒါမှမဟုတ် data file ထဲကို save လုပ်ပေးပါတယ်။
- **Export model information to XML file.** Parameter estimates နဲ့ သူတို့ရဲ့ covariances တွေကို XML (PMML) format နဲ့ သတ်မှတ်ထားတဲ့ ဖိုင်ဆီကို export လုပ်ပေးပါတယ်။ ဒီမော်ဒယ်ဖိုင်ကို အမှတ်ပေးခန့်မှန်းဖို့ (scoring purposes) တခြား data ဖိုင်တွေမှာ ပြန်သုံးနိုင်ပါတယ်။

Linear Regression Statistics (Linear Regression စာရင်းအင်းများ)

အောက်ပါ စာရင်းအင်း (statistics) တွေကို ရယူနိုင်ပါတယ်-

- **Regression Coefficients (မြောက်ဖော်ကိန်းများ)**။ *Estimates* က Regression coefficient B, standard error of B, standardized coefficient beta, t value for B နဲ့ t ရဲ့ two-tailed significance level တွေကို ပြသပေးပါတယ်။ *Confidence intervals* က သတ်မှတ်ထားတဲ့ confidence level နဲ့ coefficient တစ်ခုစီအတွက် confidence intervals တွေကို

ပြသပေးပါတယ်။ *Covariance matrix* ကတော့ off-diagonal တွေမှာ covariances နဲ့ diagonal တွေမှာ variances တွေပါဝင်တဲ့ variance-covariance matrix ကို ပြသပေးပါတယ်။

- **Model fit (မော်ဒယ်အံဝင်ခွင်ကျဖြစ်မှု)။** မော်ဒယ်ထဲ ထည့်သွင်းထားတဲ့/ဖယ်ရှားလိုက်တဲ့ variables တွေကို စာရင်းပြပေးပြီး၊ multiple R, R^2, adjusted R^2, standard error of the estimate နဲ့ ANOVA ဇယားတို့ကို ပြသပေးပါတယ်။
- **R squared change (R^2 အပြောင်းအလဲ) ။** Independent variable တစ်ခုကို ပေါင်းထည့်/ဖယ်ရှားလိုက်ခြင်းကြောင့် R^2 statistic မှာ ဖြစ်ပေါ်လာတဲ့ ပြောင်းလဲမှုပါ။ တကယ်လို့ အဲဒီအပြောင်းအလဲက ကြီးမားနေရင်၊ အဲဒီ variable ဟာ ကောင်းမွန်တဲ့ ခန့်မှန်းကိန်းတစ်ခု ဖြစ်တယ်လို့ ဆိုလိုပါတယ်။
- **Descriptives (ဖော်ပြချက်များ)။** Analysis မှာပါဝင်တဲ့ variable တစ်ခုစီအတွက် valid cases အရေအတွက်၊ mean နဲ့ standard deviation တို့ကို ပြပေးပါတယ်။
- **Partial Correlation (တစ်စိတ်တစ်ပိုင်း ဆက်စပ်မှု) ။** အခြား variables တွေရဲ့ သက်ရောက်မှုကို ဖယ်ရှားပြီးနောက် ကျန်ရှိနေတဲ့ variables နှစ်ခုကြားက ဆက်စပ်မှုပါ။
- **Part Correlation (အပိုင်းလိုက် ဆက်စပ်မှု) ။** မော်ဒယ်ထဲက အခြား independent variables တွေရဲ့ linear သက်ရောက်မှုတွေကို ဖယ်ရှားပြီးနောက်၊ dependent နဲ့ independent variable ကြားက ဆက်စပ်မှုပါ။ ၎င်းကို semipartial correlation လို့လည်း ခေါ်ပါတယ်။
- **Collinearity diagnostics.** Independent variable တစ်ခုက တခြား independent variables တွေရဲ့ linear function ဖြစ်နေတဲ့ မလိုလားအပ်တဲ့ အခြေအနေ (multicollinearity) ကို စစ်ဆေးပေးတာပါ။
- **Residuals (အကြွင်းအကျန်များ)။** Residuals တွေရဲ့ serial correlation အတွက် Durbin-Watson test နဲ့ ရွေးချယ်မှုစံနှုန်း (selection criterion) နဲ့ ကိုက်ညီတဲ့ cases တွေအတွက် casewise diagnostic information တွေကို ပြသပေးပါတယ်။

Linear Regression Options (ရွေးချယ်စရာများ)

အောက်ပါ ရွေးချယ်စရာတွေကို ရရှိနိုင်ပါတယ်-

- **Stepping Method Criteria.** Forward, backward, ဒါမှမဟုတ် stepwise variable selection နည်းလမ်းတစ်ခုခုကို သတ်မှတ်ထားတဲ့အခါမှာ အကျုံးဝင်ပါတယ်။

- ***Use Probability of F.*** F value ရဲ့ significance level က Entry value ထက်ငယ်ရင် မော်ဒယ်ထဲ ထည့်ပြီး၊ Removal value ထက်ကြီးရင် ဖယ်ရှားပါတယ်။
- ***Use F Value.*** F value က Entry value ထက်ကြီးရင် ထည့်ပြီး၊ Removal value ထက်ငယ်ရင် ဖယ်ရှားပါတယ်။

- **Include constant in equation.** ပုံမှန်အားဖြင့် regression equation မှာ constant term တစ်ခု ပါဝင်ပါတယ်။ ဒါကို အမှန်ခြစ်ဖြုတ်လိုက်ရင်တော့ regression ကို origin (သုညမှတ်) ကနေ ဖြတ်သွားစေမှာ ဖြစ်ပါတယ်။
- **Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)။** အောက်ပါတို့ထဲက တစ်ခုကို ရွေးချယ်နိုင်ပါတယ်-
  - ***Exclude cases listwise.*** Variables အားလုံးအတွက် မှန်ကန်တဲ့ (valid) တန်ဖိုးတွေရှိတဲ့ cases တွေကိုပဲ analysis ထဲ ထည့်သုံးပါတယ်။
  - ***Exclude cases pairwise.*** Correlation coefficient တွက်ချက်ဖို့ရာ ဆက်စပ်မှုတွက်မယ့် variable စုံတွဲအတွက် data ပြည့်စုံတဲ့ cases တွေကိုပဲ သုံးပါတယ်။
  - ***Replace with mean.*** တွက်ချက်မှုအားလုံးအတွက် cases အားလုံးကို အသုံးပြုပြီး၊ missing ဖြစ်နေတဲ့ နေရာတွေမှာ အဲဒီ variable ရဲ့ mean တန်ဖိုးကို အစားထိုးပြီး တွက်ပါတယ်။

REGRESSION Command Additional Features (REGRESSION Command ၏ အခြား လုပ်ဆောင်ချက်များ)

Command syntax language ကိုအသုံးပြုပြီး အောက်ပါတို့ကိုလည်း လုပ်ဆောင်နိုင်ပါသေးတယ်-

- Regression analysis ရယူဖို့အတွက် raw data တွေနေရာမှာ အစားထိုးဖို့ correlation matrix တစ်ခုကို ရေးတာ ဒါမှမဟုတ် matrix တစ်ခုကို ဖတ်တာ (MATRIX subcommand ကို သုံးပြီး) ။
- Tolerance levels တွေ သတ်မှတ်တာ (CRITERIA subcommand ကို သုံးပြီး) ။
- တူညီတဲ့ သို့မဟုတ် မတူညီတဲ့ dependent variables တွေအတွက် multiple models တွေ ရယူတာ (METHOD နှင့် DEPENDENT subcommands များကို သုံးပြီး)။
- အခြားထပ်ဆောင်း statistics တွေ ရယူတာ (DESCRIPTIVES နှင့် STATISTICS subcommands များကို သုံးပြီး)။

(ပြီးပြည့်စုံတဲ့ syntax အချက်အလက်တွေအတွက် Command Syntax Reference ကို ကြည့်ရှုနိုင်ပါတယ်။)

Ordinal Regression (အစဉ်လိုက်ဖြစ်နေသော အမျိုးအစားခွဲ ခန့်မှန်းခြင်း)

Ordinal Regression ဟာ predictors (ခန့်မှန်းကိန်းများ) အစုအဝေးတစ်ခုအပေါ် မူတည်ပြီး polytomous ordinal response (အမျိုးအစားများစွာပါဝင်ပြီး အစဉ်လိုက်ဖြစ်နေတဲ့ တုံ့ပြန်မှု) ရဲ့ မှီခိုနေမှုကို မော်ဒယ်တည်ဆောက်နိုင်စေပါတယ်။ အဲဒီ predictors တွေက factors တွေ ဒါမှမဟုတ် covariates တွေ ဖြစ်နိုင်ပါတယ်။

Standard linear regression ဟာ response (dependent) variable နဲ့ predictor (independent) variables တွေရဲ့ weighted ပေါင်းစပ်မှုကြားက sum-of-squared differences တွေကို အနည်းဆုံးဖြစ်အောင် လုပ်ဆောင်ပေးတာပါ။ ဒါပေမယ့် Ordinal variables တွေအတွက်တော့ categories တွေရဲ့ ကွာခြားချက်က ပုံသေမဟုတ်တဲ့အတွက် linear regression ကို သုံးဖို့ အမြဲတမ်း မသင့်လျော်ပါဘူး။

Example (ဥပမာ)။

လူနာတွေက ဆေးပမာဏအပေါ် တုံ့ပြန်မှုကို လေ့လာဖို့ Ordinal Regression ကို သုံးနိုင်ပါတယ်။ ဖြစ်နိုင်ချေရှိတဲ့ တုံ့ပြန်မှုတွေကို ဘာမှမဖြစ်ခြင်း (none)၊ အနည်းငယ် (mild)၊ အလယ်အလတ် (moderate)၊ ဒါမှမဟုတ် ပြင်းထန် (severe) ဆိုပြီး ခွဲခြားနိုင်ပါတယ်။ အနည်းငယ်နဲ့ အလယ်အလတ်ကြား ကွာခြားချက်က အလယ်အလတ်နဲ့ ပြင်းထန်ကြား ကွာခြားချက်ထက် ကြီးချင်လည်းကြီးမယ်၊ ငယ်ချင်လည်း ငယ်နိုင်တာမို့လို့ ဒီလိုနေရာမျိုးမှာ သုံးတာပါ။

Statistics and plots (စာရင်းအင်းများနှင့် ပုံပြဇယားများ)

Observed နဲ့ expected frequencies၊ cumulative frequencies၊ Pearson residuals၊ observed နဲ့ expected probabilities၊ parameter estimates ရဲ့ asymptotic correlation နဲ့ covariance matrices၊ Pearson's chi-square နဲ့ likelihood-ratio chi-square၊ goodness-of-fit statistics၊ iteration history၊ test of parallel lines assumption၊ parameter estimates၊ standard errors၊ confidence intervals တွေအပြင် Cox and Snell's, Nagelkerke's နဲ့ McFadden's R^2 statistics တွေကို ရရှိနိုင်ပါတယ်။

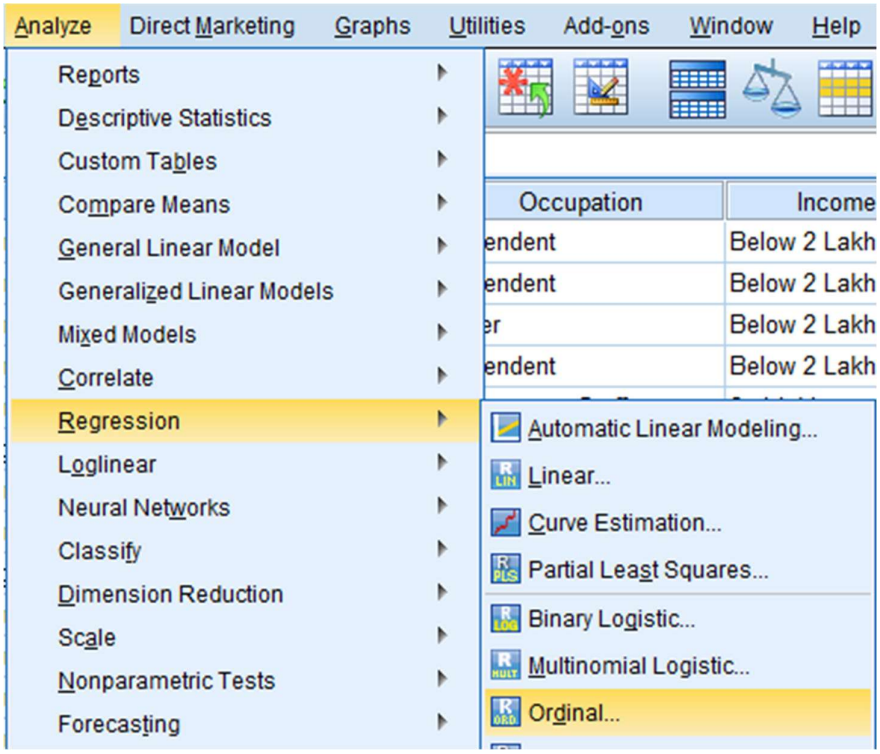
Ordinal Regression Data Considerations (Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ)

- **Data:** Dependent variable ဟာ ordinal (အစဉ်လိုက်) ဖြစ်မယ်လို့ ယူဆထားပြီး numeric သို့မဟုတ် string ဖြစ်နိုင်ပါတယ်။ Dependent variable ရဲ့ တန်ဖိုးတွေကို အငယ်မှအကြီး စီစဉ်ခြင်းဖြင့် အစဉ်လိုက်ဖြစ်မှုကို သတ်မှတ်ပါတယ်။ Factor variables တွေကိုတော့ categorical လို့ ယူဆပြီး၊ Covariate variables တွေကတော့ numeric (ဂဏန်း) ဖြစ်ရပါမယ်။
- **Assumptions (ယူဆချက်များ):** Response variable တစ်ခုတည်းကိုသာ ခွင့်ပြုပြီး မဖြစ်မနေ သတ်မှတ်ပေးရပါမယ်။ Independent variables တွေတစ်လျှောက် ကွဲပြားတဲ့

pattern တစ်ခုချင်းစီအတွက်၊ responses တွေကို အချင်းချင်း မမှီခိုတဲ့ (independent) multinomial variables တွေလို့ ယူဆထားပါတယ်။

Obtaining an Ordinal Regression (Ordinal Regression ရယူအသုံးပြုရန် အဆင့်များ)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Regression > Ordinal...



၂။ Dependent variable တစ်ခုကို ရွေးချယ်ပါ။

၃။ OK ကို နှိပ်ပါ။

Ordinal Regression Options (Ordinal Regression ဆိုင်ရာ ရွေးချယ်စရာများ)

Options dialog box ကနေတစ်ဆင့် တွက်ချက်မှု အဆင့်ဆင့် (iterative estimation algorithm) မှာ သုံးမယ့် parameters တွေကို ချိန်ညှိတာ၊ သင့်ရဲ့ parameter ခန့်မှန်းချက်တွေအတွက် ယုံကြည်မှု အတိုင်းအတာ (level of confidence) ကို ရွေးချယ်တာ နဲ့၊ link function တစ်ခုကို ရွေးချယ်တာတွေ လုပ်ဆောင်နိုင်ပါတယ်။

Iterations (တွက်ချက်မှု အဆင့်ဆင့်လုပ်ဆောင်ခြင်း) ။ တွက်ချက်မှု အဆင့်ဆင့်လုပ်ဆောင်တဲ့ iterative algorithm ကို သင်စိတ်ကြိုက် ပြောင်းလဲသတ်မှတ်နိုင်ပါတယ်။

- **Maximum iterations (အများဆုံး အဆင့်အရေအတွက်)** ။ အနှုတ်လက္ခဏာမဟုတ်တဲ့ ကိန်းပြည့် (non-negative integer) တစ်ခုကို သတ်မှတ်ပေးပါ။ တကယ်လို့ 0 လို့ သတ်မှတ်ခဲ့ရင်၊ procedure က ကနဦး ခန့်မှန်းချက် (initial estimates) တွေကိုပဲ ပြန်ထုတ်ပေးမှာ ဖြစ်ပါတယ်။
- **Maximum step-halving**။ အပေါင်းကိန်းပြည့် (positive integer) တစ်ခုကို သတ်မှတ်ပေးပါ။
- **Log-likelihood convergence**။ Log-likelihood ရဲ့ ပကတိ (absolute) ဒါမှမဟုတ် နှိုင်းရ (relative) ပြောင်းလဲမှုဟာ ဒီသတ်မှတ်ပေးလိုက်တဲ့ တန်ဖိုးထက် ငယ်သွားပြီဆိုရင် algorithm က ရပ်တန့်သွားမှာ ဖြစ်ပါတယ်။ 0 လို့ သတ်မှတ်ထားရင်တော့ ဒီသတ်မှတ်ချက်ကို အသုံးမပြုပါဘူး။
- **Parameter convergence**။ Parameter ခန့်မှန်းချက် တစ်ခုစီတိုင်းရဲ့ ပကတိ ဒါမှမဟုတ် နှိုင်းရ ပြောင်းလဲမှုဟာ ဒီသတ်မှတ်ပေးလိုက်တဲ့ တန်ဖိုးထက် ငယ်သွားပြီဆိုရင် algorithm က ရပ်တန့်သွားမှာပါ။ 0 လို့ သတ်မှတ်ထားရင်တော့ ဒီသတ်မှတ်ချက်ကို အသုံးမပြုပါဘူး။

Confidence interval (ယုံကြည်မှု အတိုင်းအတာ) ။ 0 နဲ့ ညီတာ ဒါမှမဟုတ် 0 ထက်ကြီးပြီး၊ 100 ထက်ငယ်တဲ့ တန်ဖိုးတစ်ခုကို သတ်မှတ်ပေးပါ။

Delta။ သုညဖြစ်နေတဲ့ cell frequencies တွေအတွက် ပေါင်းထည့်ပေးမယ့် တန်ဖိုးဖြစ်ပါတယ်။ 1 ထက်ငယ်တဲ့ အနှုတ်မဟုတ်သော တန်ဖိုးတစ်ခုကို သတ်မှတ်ပေးပါ။

Singularity tolerance။ အချင်းချင်း အရမ်းမှီခိုနေတဲ့ (highly dependent) predictors တွေကို စစ်ဆေးဖို့ အသုံးပြုပါတယ်။ ရွေးချယ်စရာ စာရင်းထဲကနေ တန်ဖိုးတစ်ခုကို ရွေးချယ်ပေးပါ။

Link function။ Link function ဆိုတာ မော်ဒယ်ကို ခန့်မှန်းတွက်ချက်နိုင်ဖို့အတွက် cumulative probabilities (စုဆောင်းဖြစ်တန်စွမ်းများ) ကို ပုံစံပြောင်းလဲပေးခြင်း (transformation) ဖြစ်ပါတယ်။ အောက်ဖော်ပြပါ link functions ၅ မျိုးကို ရွေးချယ်အသုံးပြုနိုင်ပါတယ်-

- **Logit**။ f(x)=log(x/(1–x)) ဖြစ်ပါတယ်။ ပုံမှန်အားဖြင့် categories တွေ ညီညီမျှမျှ ဖြန့်ကျက် (evenly distributed) ဖြစ်နေတဲ့ အခါမျိုးမှာ သုံးလေ့ရှိပါတယ်။
- **Complementary log-log**။ f(x)=log(–log(1–x)) ဖြစ်ပါတယ်။ ပုံမှန်အားဖြင့် ပိုမြင့်တဲ့ categories တွေ ဖြစ်နိုင်ချေ ပိုများတဲ့ (more probable) အခါမျိုးမှာ သုံးလေ့ရှိပါတယ်။
- **Negative log-log**။ f(x)=–log(–log(x)) ဖြစ်ပါတယ်။ ပုံမှန်အားဖြင့် ပိုနိမ့်တဲ့ categories တွေ ဖြစ်နိုင်ချေ ပိုများတဲ့ အခါမျိုးမှာ သုံးလေ့ရှိပါတယ်။
- **Probit**။ f(x)=Φ–1(x) ဖြစ်ပါတယ်။ ပုံမှန်အားဖြင့် မမြင်ရတဲ့ လျှို့ဝှက်ပြောင်းလဲကိန်း (latent variable) ဟာ normal distribution (ပုံမှန်ဖြန့်ကျက်မှု) ရှိနေတဲ့ အခါမျိုးမှာ သုံးလေ့ရှိပါတယ်။

- **Cauchit (inverse Cauchy)** ။ f(x)=tan(π(x–0.5)) ဖြစ်ပါတယ်။ ပုံမှန်အားဖြင့် latent variable မှာ လွန်ကဲတဲ့တန်ဖိုးတွေ (extreme values) အများကြီး ရှိနေတဲ့ အခါမျိုးမှာ သုံးလေ့ရှိပါတယ်။

Ordinal Regression Output (ရလဒ် ထုတ်ယူခြင်း)

Output dialog box ကနေတစ်ဆင့် Viewer မှာ ပြသဖို့အတွက် ဇယား (tables) တွေ ထုတ်ပေးနိုင်သလို၊ ပြောင်းလဲကိန်း (variables) တွေကိုလည်း လက်ရှိအသုံးပြုနေတဲ့ working file ထဲကို သိမ်းဆည်း (save) နိုင်ပါတယ်။

Display (ပြသရန်)

အောက်ပါ အချက်အလက်တွေအတွက် ဇယားတွေ ထုတ်ပေးပါတယ်-

- **Print iteration history for every n step(s) (အဆင့် n ခုတိုင်းအတွက် တွက်ချက်မှုမှတ်တမ်းကို ပြသခြင်း)** ။ သတ်မှတ်ပေးထားတဲ့ အကြိမ်အရေအတွက် (frequency) အတိုင်း log-likelihood နဲ့ parameter ခန့်မှန်းချက်တွေကို ပြသပေးပါတယ်။ ပထမဆုံးအကြိမ်နဲ့ နောက်ဆုံးအကြိမ် တွက်ချက်မှု (iterations) တွေကိုတော့ အမြဲတမ်း ပြသပေးပါတယ်။

- **Goodness of fit statistics (အံဝင်ခွင်ကျဖြစ်မှုဆိုင်ရာ စာရင်းအင်းများ)** ။ Pearson နဲ့ likelihood-ratio chi-square စာရင်းအင်းတွေ ဖြစ်ပါတယ်။ သူတို့ကို variable စာရင်းမှာ သတ်မှတ်ထားတဲ့ အမျိုးအစားခွဲခြားမှု (classification) အပေါ် အခြေခံပြီး တွက်ချက်ထားတာပါ။

- **Summary statistics (အကျဉ်းချုပ် စာရင်းအင်းများ)** ။ Cox and Snell's, Nagelkerke's, နဲ့ McFadden's R^2 စာရင်းအင်းတွေ ပါဝင်ပါတယ်။

- **Parameter estimates (ပါရာမီတာ ခန့်မှန်းချက်များ)** ။ Parameter ခန့်မှန်းချက်တွေ၊ စံအမှား (standard errors) တွေနဲ့ ယုံကြည်မှု အတိုင်းအတာ (confidence intervals) တွေ ပါဝင်ပါတယ်။

- **Asymptotic correlation of parameter estimates (ပါရာမီတာ ခန့်မှန်းချက်များ၏ Asymptotic ဆက်စပ်မှု)** ။ Parameter ခန့်မှန်းချက်တွေရဲ့ correlations တွေကို matrix ဇယားအနေနဲ့ ပြသပေးတာပါ။

- **Asymptotic covariance of parameter estimates (ပါရာမီတာ ခန့်မှန်းချက်များ၏ Asymptotic ကွဲလွဲမှု)** ။ Parameter ခန့်မှန်းချက်တွေရဲ့ covariances တွေကို matrix ဇယားအနေနဲ့ ပြသပေးတာပါ။

- **Cell information (ဆဲလ် အချက်အလက်များ)။** Covariate ပုံစံတစ်ခုချင်းစီအလိုက် တုံ့ပြန်မှု အမျိုးအစား (response category) တစ်ခုစီအတွက် လေ့လာတွေ့ရှိရတဲ့ (observed) နဲ့ မျှော်မှန်းထားတဲ့ (expected) frequencies တွေ၊ cumulative frequencies တွေ၊ frequencies နဲ့ cumulative frequencies တွေအတွက် Pearson residuals တွေ၊ observed နဲ့ expected ဖြစ်တန်စွမ်း (probabilities) တွေအပြင် observed နဲ့ expected cumulative probabilities တွေကိုပါ ပြသပေးပါတယ်,။ မှတ်သားထားရမယ့် အချက်ကတော့၊ covariate ပုံစံ (patterns) တွေ အများကြီးပါဝင်တဲ့ မော်ဒယ်တွေ (ဥပမာ - continuous covariates ပါဝင်တဲ့ မော်ဒယ်တွေ) အတွက်ဆိုရင် ဒီ option က အရမ်းကြီးမားပြီး ကိုင်တွယ်ရခက်တဲ့ ဇယားကြီးတစ်ခုကို ထုတ်ပေးနိုင်ပါတယ်။
- **Test of parallel lines (အပြိုင်မျဉ်းများ စမ်းသပ်ခြင်း) ။** Location parameters တွေဟာ dependent variable ရဲ့ အဆင့်တွေတစ်လျှောက် တူညီမှုရှိတယ်ဆိုတဲ့ ယူဆချက် (hypothesis) ကို စမ်းသပ်ပေးတာပါ။ ဒီ option ကို location-only မော်ဒယ်အတွက်သာ အသုံးပြုနိုင်ပါတယ်။

Saved Variables (သိမ်းဆည်းမည့် ပြောင်းလဲကိန်းများ)

အောက်ဖော်ပြပါ variables တွေကို working file ထဲမှာ သိမ်းဆည်းပေးပါတယ်-

- **Estimated response probabilities (ခန့်မှန်း တုံ့ပြန်မှု ဖြစ်တန်စွမ်းများ) ။** Factor/covariate ပုံစံတစ်ခုကို တုံ့ပြန်မှု အမျိုးအစား (response categories) တွေထဲကို ခွဲခြားသတ်မှတ်ဖို့အတွက် မော်ဒယ်ကနေ ခန့်မှန်းပေးတဲ့ ဖြစ်တန်စွမ်းတွေ ဖြစ်ပါတယ်။ Response categories အရေအတွက် ဘယ်လောက်ရှိသလဲ၊ အဲဒီအရေအတွက်အတိုင်း probabilities တွေ ရရှိမှာပါ။
- **Predicted category (ခန့်မှန်း အမျိုးအစား) ။** သီးခြား factor/covariate ပုံစံတစ်ခုအတွက် အများဆုံး (maximum) ခန့်မှန်းဖြစ်တန်စွမ်းရှိတဲ့ တုံ့ပြန်မှု အမျိုးအစား ဖြစ်ပါတယ်။
- **Predicted category probability (ခန့်မှန်း အမျိုးအစား၏ ဖြစ်တန်စွမ်း) ။** Factor/covariate ပုံစံတစ်ခုကို predicted category ထဲ ထည့်သွင်းသတ်မှတ်နိုင်ခြေ ခန့်မှန်းဖြစ်တန်စွမ်း ဖြစ်ပါတယ်။ ဒီ ဖြစ်တန်စွမ်းဟာ factor/covariate ပုံစံတစ်ခုရဲ့ ခန့်မှန်းဖြစ်တန်စွမ်းတွေထဲမှာ အများဆုံးတန်ဖိုး (maximum) လည်း ဖြစ်ပါတယ်။
- **Actual category probability (အမှန်တကယ် အမျိုးအစား၏ ဖြစ်တန်စွမ်း) ။** Factor/covariate ပုံစံတစ်ခုကို အမှန်တကယ် အမျိုးအစား (actual category) ထဲ ထည့်သွင်းသတ်မှတ်နိုင်ခြေ ခန့်မှန်းဖြစ်တန်စွမ်း ဖြစ်ပါတယ်။

Print Log-Likelihood (Log-Likelihood ကို ပြသရန်)

Log-likelihood ရဲ့ ပြသမှုကို ထိန်းချုပ်ပေးပါတယ်။ Multinomial constant ကို ထည့်သွင်းထားမယ်ဆိုရင်တော့ likelihood ရဲ့ အပြည့်အဝတန်ဖိုး (full value) ကို ရရှိစေမှာပါ။ တကယ်လို့ သင်က သင့်ရဲ့ ရလဒ်တွေကို constant မပါဝင်တဲ့ အခြားထုတ်ကုန် (products) တွေနဲ့ နှိုင်းယှဉ်ချင်တယ် ဆိုရင်တော့၊ ၎င်းကို မထည့်သွင်းဘဲ (exclude) ဖယ်ထုတ်ဖို့ ရွေးချယ်နိုင်ပါတယ်။

Ordinal Regression Location Model (Ordinal Regression လိုကေးရှင်း မော်ဒယ်)

Location dialog box ကနေတစ်ဆင့် သင့်ရဲ့ analysis (ခွဲခြမ်းစိတ်ဖြာမှု) အတွက် location model ကို သတ်မှတ်ပေးနိုင်ပါတယ်။

- **Specify model (မော်ဒယ်သတ်မှတ်ခြင်း)။** Main-effects မော်ဒယ်တစ်ခုမှာ covariate နဲ့ factor တွေရဲ့ အဓိကသက်ရောက်မှု (main effects) တွေ ပါဝင်ပေမယ့်၊ အချင်းချင်းသက်ရောက်မှု (interaction effects) တွေတော့ မပါဝင်ပါဘူး။ Factor interactions သို့မဟုတ် covariate interactions အုပ်စုကွဲလေးတွေကို သီးခြားသတ်မှတ်ဖို့အတွက် custom model (စိတ်ကြိုက်မော်ဒယ်) တစ်ခုကို သင် ဖန်တီးနိုင်ပါတယ်။
- **Factors/covariates။** Factors နဲ့ covariates တွေကို စာရင်းပြထားပါတယ်။
- **Location model (တည်နေရာမော်ဒယ်)။** ဒီမော်ဒယ်က သင်ရွေးချယ်လိုက်တဲ့ main effects နဲ့ interaction effects တွေအပေါ် မူတည်ပါတယ်။

ရွေးချယ်ထားတဲ့ factors နဲ့ covariates တွေအတွက် အောက်ပါတို့ကို ရွေးချယ်နိုင်ပါတယ်-

- **Interaction (အချင်းချင်း သက်ရောက်မှု) ။** ရွေးချယ်ထားတဲ့ variables အားလုံးရဲ့ အမြင့်ဆုံးအဆင့် (highest-level) interaction term ကို ဖန်တီးပေးပါတယ်။ ဒါက ပုံမှန် (default) အပြင်အဆင်ပါ။
- **Main effects (အဓိက သက်ရောက်မှုများ) ။** ရွေးချယ်ထားတဲ့ variable တစ်ခုချင်းစီအတွက် main-effects term တစ်ခုစီ ဖန်တီးပေးပါတယ်။
- **All 2-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၂-ခုတွဲ (two-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။
- **All 3-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၃-ခုတွဲ (three-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။

- **All 4-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၄-ခုတွဲ (four-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။
- **All 5-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၅-ခုတွဲ (five-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။

Build Terms and Custom Terms (သတ်မှတ်ချက်များ တည်ဆောက်ခြင်း)

- **Build terms (သတ်မှတ်ချက်များ တည်ဆောက်ခြင်း) ။** ရွေးချယ်ထားတဲ့ factors နဲ့ covariates အစုအဝေးတစ်ခုရဲ့ ပေါင်းစပ်မှုအားလုံးအတွက် သီးခြားအမျိုးအစားတစ်ခုခုရဲ့ (ဥပမာ- main effects လိုမျိုး) non-nested terms တွေကို ထည့်သွင်းချင်တဲ့အခါ ဒီရွေးချယ်မှုကို အသုံးပြုပါ။
- **Build custom terms (စိတ်ကြိုက်သတ်မှတ်ချက်များ တည်ဆောက်ခြင်း)** ထပ်ဆင့်ပါဝင်နေတဲ့ (nested) terms တွေကို ထည့်သွင်းချင်တဲ့အခါ သို့မဟုတ် term တစ်ခုခုကို variable တစ်ခုချင်းစီအလိုက် တိတိကျကျ တည်ဆောက်ချင်တဲ့အခါ ဒီရွေးချယ်မှုကို သုံးပါ။ Nested term တစ်ခု တည်ဆောက်ရာမှာ အောက်ပါ အဆင့်တွေ ပါဝင်ပါတယ်-

Ordinal Regression Scale Model (Ordinal Regression စကေး မော်ဒယ်)

Scale dialog box ကနေတစ်ဆင့် သင့်ရဲ့ analysis အတွက် scale model ကို သတ်မှတ်ပေးနိုင်ပါတယ်။

- **Factors/covariates။** Factors နဲ့ covariates တွေကို စာရင်းပြထားပါတယ်။
- **Scale model (စကေးမော်ဒယ်)။** ဒီမော်ဒယ်က သင်ရွေးချယ်လိုက်တဲ့ main နဲ့ interaction effects တွေအပေါ် မူတည်ပါတယ်။

ရွေးချယ်ထားတဲ့ factors နဲ့ covariates တွေအတွက် အောက်ပါတို့ကို ရွေးချယ်နိုင်ပါတယ်-

- **Interaction (အချင်းချင်း သက်ရောက်မှု) ။** ရွေးချယ်ထားတဲ့ variables အားလုံးရဲ့ အမြင့်ဆုံးအဆင့် (highest-level) interaction term ကို ဖန်တီးပေးပါတယ်။ ဒါက ပုံမှန် (default) အပြင်အဆင်ပါ။
- **Main effects (အဓိက သက်ရောက်မှုများ) ။** ရွေးချယ်ထားတဲ့ variable တစ်ခုချင်းစီအတွက် main-effects term တစ်ခုစီ ဖန်တီးပေးပါတယ်။
- **All 2-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၂-ခုတွဲ (two-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။

- **All 3-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၃-ခုတွဲ (three-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။
- **All 4-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၄-ခုတွဲ (four-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။
- **All 5-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၅-ခုတွဲ (five-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။

Build Terms and Custom Terms (သတ်မှတ်ချက်များ တည်ဆောက်ခြင်း)

- **Build terms (သတ်မှတ်ချက်များ တည်ဆောက်ခြင်း) ။** ရွေးချယ်ထားတဲ့ factors နဲ့ covariates အစုအဝေးတစ်ခုရဲ့ ပေါင်းစပ်မှုအားလုံးအတွက် သီးခြားအမျိုးအစားတစ်ခုခုရဲ့ (ဥပမာ- main effects လိုမျိုး) non-nested terms တွေကို ထည့်သွင်းချင်တဲ့အခါ ဒီရွေးချယ်မှုကို အသုံးပြုပါ။
- **Build custom terms (စိတ်ကြိုက်သတ်မှတ်ချက်များ တည်ဆောက်ခြင်း) ။** ထပ်ဆင့်ပါဝင်နေတဲ့ (nested) terms တွေကို ထည့်သွင်းချင်တဲ့အခါ သို့မဟုတ် term တစ်ခုခုကို variable တစ်ခုချင်းစီအလိုက် တိတိကျကျ တည်ဆောက်ချင်တဲ့အခါ ဒီရွေးချယ်မှုကို သုံးပါ။ Nested term တစ်ခု တည်ဆောက်ရာမှာ အောက်ပါ အဆင့်တွေ ပါဝင်ပါတယ်-

(မှတ်ချက် - ဤနေရာတွင် အောက်ဆက်လက်လုပ်ဆောင်ရမည့် အဆင့်များကို မူရင်းစာသားနှင့် ချန်လှပ်ထားပါသည်။)

Curve Estimation (မျဉ်းကွေးပုံစံ ခန့်မှန်းခြင်း)

Curve Estimation လုပ်ဆောင်ချက်ဟာ မတူညီတဲ့ curve estimation regression မော်ဒယ် ၁၁ မျိုးအတွက် regression စာရင်းအင်းတွေနဲ့ ဆက်စပ်တဲ့ plots (ပုံပြဇယားများ) ကို ထုတ်ပေးပါတယ်။ Dependent variable တစ်ခုချင်းစီအတွက် သီးခြားမော်ဒယ်တစ်ခုစီကို ဖန်တီးပေးတာ ဖြစ်ပါတယ်။ ဒါ့အပြင် predicted values (ခန့်မှန်းတန်ဖိုးများ)၊ residuals (ကွာဟချက်များ) နဲ့ prediction intervals တွေကိုလည်း variable အသစ်တွေအနေနဲ့ သိမ်းဆည်း (save) နိုင်ပါတယ်။

Example (ဥပမာ)။ အင်တာနက် ဝန်ဆောင်မှုပေးတဲ့ ကုမ္ပဏီတစ်ခုက သူတို့ရဲ့ ကွန်ရက်ပေါ်မှာ ဗိုင်းရပ်စ် ကူးစက်ခံရတဲ့ အီးမေးလ် ရာခိုင်နှုန်းကို အချိန်နဲ့အမျှ စောင့်ကြည့်လေ့လာတယ် ဆိုပါစို့။ Scatterplot (အစက်ချဇယား) ဆွဲကြည့်တဲ့အခါ သူတို့ရဲ့ ဆက်စပ်မှုက မျဉ်းဖြောင့် (nonlinear) မဟုတ်ဘူးဆိုတာကို တွေ့ရပါတယ်။ ဒီအခါမျိုးမှာ သင့်အနေနဲ့ quadratic ဒါမှမဟုတ် cubic မော်ဒယ်ကို အသုံးပြုပြီး အချက်အလက်တွေနဲ့ အံဝင်ခွင်ကျဖြစ်အောင် (fit) တည်ဆောက်နိုင်သလို၊ ယူဆချက် (assumptions) တွေ မှန်ကန်မှုရှိမရှိနဲ့ မော်ဒယ်ရဲ့ goodness of fit (ဘယ်လောက် အံဝင်ခွင်ကျဖြစ်သလဲ) ကိုပါ စစ်ဆေးကြည့်နိုင်ပါတယ်။

Statistics (စာရင်းအင်းများ)။ မော်ဒယ်တစ်ခုချင်းစီအတွက် regression coefficients၊ multiple R, R^2, adjusted R^2၊ standard error of the estimate၊ analysis-of-variance table (ANOVA ဇယား) ၊ predicted values၊ residuals နဲ့ prediction intervals တွေကို ရရှိနိုင်ပါတယ်။ မော်ဒယ်အမျိုးအစားတွေ အနေနဲ့ကတော့ linear, logarithmic, inverse, quadratic, cubic, power, compound, S-curve, logistic, growth နဲ့ exponential တို့ ပါဝင်ပါတယ်။

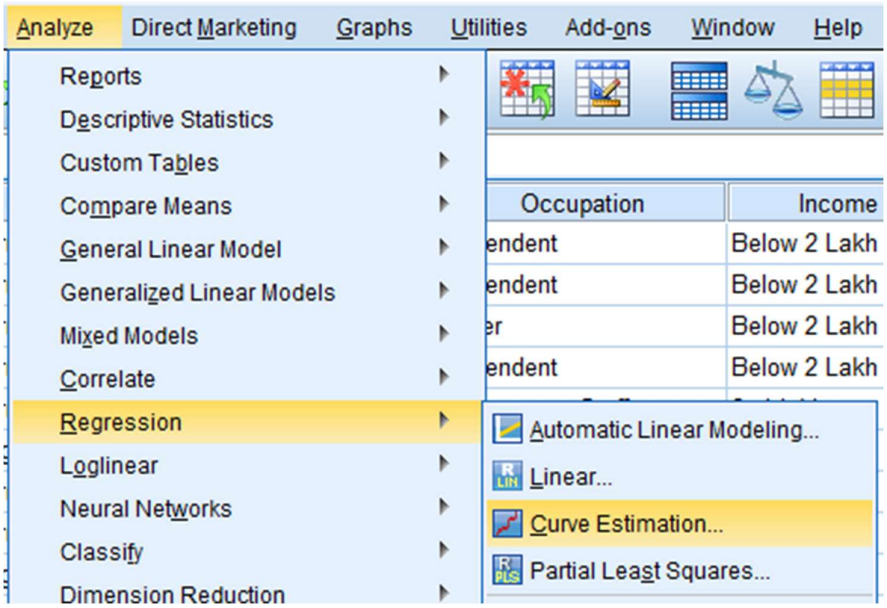
Curve Estimation Data Considerations (Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ)

- **Data (အချက်အလက်)။** Dependent နဲ့ independent variables တွေဟာ အရေအတွက်ပြ (quantitative) ဖြစ်ရပါမယ်။ တကယ်လို့ သင်က independent variable နေရာမှာ variable တစ်ခုခုကို မရွေးချယ်ဘဲ active dataset ထဲက 'Time' ကိုသာ ရွေးချယ်ခဲ့မယ်ဆိုရင်၊ Curve Estimation procedure က cases တွေကြားမှာ တူညီတဲ့ အချိန်အကွာအဝေး (uniform length of time) ရှိတဲ့ time variable တစ်ခုကို အလိုအလျောက် ဖန်တီးပေးမှာ ဖြစ်ပါတယ်။ Time ကို ရွေးချယ်ခဲ့ရင်၊ dependent variable ဟာ time-series (အချိန်စဉ်ဆက်) အတိုင်းအတာတစ်ခု ဖြစ်သင့်ပါတယ်။ Time-series analysis လုပ်ဖို့အတွက် case (အတန်း) တစ်ခုစီဟာ မတူညီတဲ့ အချိန်တစ်ခုစီက လေ့လာတွေ့ရှိချက်တွေကို ကိုယ်စားပြုရမယ့်အပြင်၊ cases တွေကြားက အချိန်အကွာအဝေးဟာ တူညီညီမျှ (uniform) ဖြစ်နေတဲ့ data file ဖွဲ့စည်းပုံမျိုး လိုအပ်ပါတယ်။

- **Assumptions (ယူဆချက်များ)။** Independent နဲ့ dependent variables တွေ ဘယ်လိုဆက်စပ်နေလဲ (ဥပမာ- မျဉ်းဖြောင့် linearly လား၊ exponentially လား စသဖြင့်) သိနိုင်ဖို့ သင့်ရဲ့ data တွေကို ပုံပြဇယား (graphically) အရင်ဆုံး စစ်ဆေးသင့်ပါတယ်။ ကောင်းမွန်တဲ့ မော်ဒယ်တစ်ခုရဲ့ residuals တွေဟာ ကျပန်းဖြန့်ကျက် (randomly distributed) ဖြစ်နေရမယ့်အပြင် normal လည်း ဖြစ်နေရပါမယ်။ တကယ်လို့ linear model ကို အသုံးပြုမယ်ဆိုရင် အောက်ပါ ယူဆချက်တွေနဲ့ ကိုက်ညီဖို့ လိုအပ်ပါတယ် - independent variable ရဲ့ တန်ဖိုးတစ်ခုချင်းစီတိုင်းအတွက် dependent variable ရဲ့ ဖြန့်ကျက်မှု (distribution) ဟာ normal ဖြစ်ရပါမယ်။ ပြီးတော့ independent variable ရဲ့ တန်ဖိုးအားလုံးအတွက် dependent variable distribution ရဲ့ variance ဟာ ကိန်းသေ (constant) ဖြစ်သင့်ပါတယ်။ Dependent variable နဲ့ independent variable ကြားက ဆက်စပ်မှုက linear (မျဉ်းဖြောင့်) ဖြစ်ရမယ့်အပြင်၊ လေ့လာတွေ့ရှိချက် (observations) အားလုံးကလည်း အချင်းချင်း မှီခိုနေခြင်းမရှိ (independent) ဖြစ်ရပါမယ်။

To Obtain a Curve Estimation (Curve Estimation ရယူအသုံးပြုရန် အဆင့်များ)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Regression > Curve Estimation...



၂။ Dependent variable တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။ (Dependent variable တစ်ခုချင်းစီအတွက် သီးခြားမော်ဒယ်တစ်ခုစီ ထုတ်ပေးပါလိမ့်မယ်)။

၃။ Independent variable တစ်ခုကို ရွေးချယ်ပါ (active dataset ထဲက variable တစ်ခုခုကို ရွေးလို့ရသလို၊ 'Time' ကိုလည်း ရွေးချယ်နိုင်ပါတယ်)။

၄။ ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့ အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- Scatterplots တွေမှာ cases တွေကို အမည်တပ် (label) ဖို့ variable တစ်ခုကို ရွေးချယ်နိုင်ပါတယ်။ Scatterplot ပေါ်က အမှတ် (point) တစ်ခုချင်းစီအတွက် Case Label variable ရဲ့ တန်ဖိုးကို ပြသဖို့ Point Selection tool ကို အသုံးပြုနိုင်ပါတယ်။
- Predicted values၊ residuals နဲ့ prediction intervals တွေကို variables အသစ်တွေအနေနဲ့ သိမ်းဆည်း (save) ဖို့ **Save** ကို နှိပ်နိုင်ပါတယ်။

အောက်ဖော်ပြပါ ရွေးချယ်စရာ (options) တွေကိုလည်း အသုံးပြုနိုင်ပါသေးတယ်-

- **Include constant in equation (ညီမျှခြင်းတွင် ကိန်းသေ ထည့်သွင်းရန်)** ။ Regression equation ထဲမှာ ကိန်းသေ (constant term) တစ်ခုကို ခန့်မှန်းတွက်ချက်ပေးပါတယ်။ ပုံမှန်အားဖြင့် (by default) ကိန်းသေကို ထည့်သွင်းထားပါတယ်။
- **Plot models (မော်ဒယ်များအား ပုံဆွဲရန်)** ။ ရွေးချယ်ထားတဲ့ မော်ဒယ်တစ်ခုစီတိုင်းနဲ့ dependent variable တန်ဖိုးတွေကို independent variable အပေါ်မူတည်ပြီး ပုံဆွဲ (plot) ပေးပါတယ်။ Dependent variable တစ်ခုချင်းစီအတွက် သီးခြား chart တစ်ခုစီ ထုတ်ပေးမှာပါ။
- **Display ANOVA table (ANOVA ဇယား ပြသရန်)** ။ ရွေးချယ်ထားတဲ့ မော်ဒယ်တစ်ခုချင်းစီအတွက် အကျဉ်းချုပ် analysis-of-variance ဇယားကို ပြသပေးပါတယ်။

Curve Estimation Models (မော်ဒယ်အမျိုးအစားများ)

Curve estimation regression မော်ဒယ် တစ်ခု ဒါမှမဟုတ် တစ်ခုထက်ပိုပြီး သင် ရွေးချယ်နိုင်ပါတယ်။ ဘယ်မော်ဒယ်ကို သုံးရမလဲဆိုတာ ဆုံးဖြတ်ဖို့ သင့်ရဲ့ data ကို plot ဆွဲကြည့်ပါ။ တကယ်လို့ variables တွေက linear (မျဉ်းဖြောင့်) ဆက်စပ်မှုရှိပုံရရင်၊ ရိုးရှင်းတဲ့ simple linear regression မော်ဒယ်ကို အသုံးပြုပါ။ Variables တွေက linearly ဆက်စပ်မနေဘူးဆိုရင်တော့ သင့်ရဲ့ data ကို transform (ပြောင်းလဲ) ကြည့်ပါ။ Data ကို transform လုပ်တာက အလုပ်မဖြစ်ဘူးဆိုရင်တော့ ပိုပြီးရှုပ်ထွေးတဲ့ မော်ဒယ်တစ်ခု လိုအပ်လာပါပြီ။ သင့် data ရဲ့ scatterplot ကို ကြည့်ပါ၊ အဲဒီပုံစံက သင်သိတဲ့ mathematical function တစ်ခုခုနဲ့ ဆင်တူနေရင် အဲဒီလို မော်ဒယ်အမျိုးအစားမျိုးကို သင့် data နဲ့ fit လုပ်ကြည့်ပါ။ (ဥပမာ - data ပုံစံက exponential function နဲ့ ဆင်နေရင် exponential မော်ဒယ်ကို သုံးပါ)။

ရွေးချယ်နိုင်တဲ့ မော်ဒယ်တွေကတော့-

- **Linear (မျဉ်းဖြောင့်)။** Y = b0 + (b1 * t) ဖြစ်တဲ့ မော်ဒယ်ပါ။ Series တန်ဖိုးတွေကို အချိန် (time) ရဲ့ linear function အဖြစ် မော်ဒယ်တည်ဆောက်ပါတယ်။
- **Logarithmic (လော့ဂရစ်သမ်)။** Y = b0 + (b1 * ln(t)) ဖြစ်တဲ့ မော်ဒယ်ပါ။
- **Inverse (ပြောင်းပြန်)။** Y = b0 + (b1 / t) ဖြစ်တဲ့ မော်ဒယ်ပါ။
- **Quadratic (နှစ်ထပ်ကိန်း)။** Y = b0 + (b1 * t) + (b2 * t**2) ဖြစ်တဲ့ မော်ဒယ်ပါ။ ဒီမော်ဒယ်ကို ရုတ်တရက် ထိုးတက်သွားတဲ့ (takes off) ဒါမှမဟုတ် တဖြည်းဖြည်း လျော့ကျသွားတဲ့ (dampens) series တွေအတွက် သုံးနိုင်ပါတယ်။
- **Cubic (သုံးထပ်ကိန်း)။** Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3) နဲ့ သတ်မှတ်ထားတဲ့ မော်ဒယ်ပါ။
- **Power။** Y = b0 * (t**b1) ဒါမှမဟုတ် ln(Y) = ln(b0) + (b1 * ln(t)) ဖြစ်တဲ့ မော်ဒယ်ပါ။
- **Compound။** Y = b0 * (b1**t) ဒါမှမဟုတ် ln(Y) = ln(b0) + (ln(b1) * t) ဖြစ်တဲ့ မော်ဒယ်ပါ။
- **S-curve။** Y = e**(b0 + (b1/t)) ဒါမှမဟုတ် ln(Y) = b0 + (b1/t) ဖြစ်တဲ့ မော်ဒယ်ပါ။
- **Logistic။** Y = 1 / (1/u + (b0 * (b1**t))) ဒါမှမဟုတ် ln(1/y-1/u) = ln (b0) + (ln(b1) * t) ဖြစ်တဲ့ မော်ဒယ်ဖြစ်ပြီး u ဆိုတာကတော့ upper boundary value (အမြင့်ဆုံး ကန့်သတ်တန်ဖိုး) ဖြစ်ပါတယ်။ Logistic ကို ရွေးချယ်ပြီးရင်၊ regression equation မှာ အသုံးပြုမယ့် upper boundary value ကို သတ်မှတ်ပေးရပါမယ်။ ဒီတန်ဖိုးဟာ အကြီးဆုံး dependent variable ရဲ့ တန်ဖိုးထက် ကြီးတဲ့ အပေါင်းကိန်းဂဏန်း (positive number) တစ်ခု ဖြစ်ရပါမယ်။
- **Growth (တိုးတက်မှု)။** Y = e**(b0 + (b1 * t)) ဒါမှမဟုတ် ln(Y) = b0 + (b1 * t) ဖြစ်တဲ့ မော်ဒယ်ပါ။
- **Exponential။** Y = b0 * (e**(b1 * t)) ဒါမှမဟုတ် ln(Y) = ln(b0) + (b1 * t) ဖြစ်တဲ့ မော်ဒယ်ပါ။

Curve Estimation Save (Curve Estimation သိမ်းဆည်းခြင်း)

- **Save Variables (ပြောင်းလဲကိန်းများ သိမ်းဆည်းခြင်း) ။** ရွေးချယ်ထားတဲ့ မော်ဒယ်တစ်ခုချင်းစီအတွက်၊ predicted values၊ residuals (dependent variable ရဲ့ လေ့လာတွေ့ရှိရတဲ့တန်ဖိုးထဲကနေ မော်ဒယ်က ခန့်မှန်းထားတဲ့ တန်ဖိုးကို နှုတ်ထားတဲ့ရလဒ်) နဲ့ prediction intervals (အထက်နဲ့ အောက် ကန့်သတ်ချက်များ) တို့ကို

သိမ်းဆည်းနိုင်ပါတယ်။ Variable အမည်သစ်တွေနဲ့ ရှင်းလင်းဖော်ပြချက် (descriptive labels) တွေကို output window ထဲက ဇယားတစ်ခုမှာ ပြသပေးပါလိမ့်မယ်။

- **Predict Cases (အခြေအနေများ ခန့်မှန်းခြင်း)** ။ Active dataset ထဲမှာ သင်က independent variable အဖြစ် variable တစ်ခုကို မရွေးဘဲ 'Time' ကို ရွေးချယ်ထားခဲ့မယ်ဆိုရင်၊ time series အဆုံးသတ်ပြီးနောက်ပိုင်းကာလအတွက် ကြိုတင်ခန့်မှန်းမယ့် ကာလ (forecast period) ကို သတ်မှတ်ပေးနိုင်ပါတယ်။ အောက်ပါ ရွေးချယ်စရာတွေထဲက တစ်ခုကို ရွေးချယ်နိုင်ပါတယ်-
    - **Predict from estimation period through last case**။ Estimation period ထဲက cases တွေအပေါ် အခြေခံပြီး ဖိုင်ထဲမှာရှိတဲ့ cases အားလုံးအတွက် တန်ဖိုးတွေကို ခန့်မှန်းပေးပါတယ်။ Dialog box ရဲ့ အောက်ခြေမှာ ပြသထားတဲ့ Estimation period ဆိုတာကို Data menu ပေါ်က Select Cases option ရဲ့ Range subdialog box ထဲမှာ သတ်မှတ်ပေးထားတာပါ။ တကယ်လို့ estimation period ကို သတ်မှတ်မထားဘူးဆိုရင်တော့ တန်ဖိုးတွေခန့်မှန်းဖို့ cases အားလုံးကို အသုံးပြုသွားမှာ ဖြစ်ပါတယ်။
    - **Predict through**။ Estimation period ထဲက cases တွေအပေါ် အခြေခံပြီး သတ်မှတ်ထားတဲ့ ရက်စွဲ၊ အချိန်၊ ဒါမှမဟုတ် လေ့လာမှုအမှတ်စဉ် (observation number) အထိ တန်ဖိုးတွေကို ခန့်မှန်းပေးပါတယ်။ ဒီလုပ်ဆောင်ချက်ကို time series ထဲက နောက်ဆုံး case အလွန်အထိ တန်ဖိုးတွေကို ကြိုတင်ခန့်မှန်း (forecast) ဖို့အတွက် သုံးနိုင်ပါတယ်။ လက်ရှိ သတ်မှတ်ထားတဲ့ date variables တွေက ခန့်မှန်းမယ့်ကာလ အဆုံးသတ်ဖို့အတွက် ဘယ် text boxes တွေကို အသုံးပြုနိုင်မလဲဆိုတာ ဆုံးဖြတ်ပေးပါတယ်။ တကယ်လို့ date variables တွေကို သတ်မှတ်မထားဘူးဆိုရင်တော့ အဆုံးသတ်မယ့် observation (case) နံပါတ်ကို သင်ကိုယ်တိုင် သတ်မှတ်ပေးနိုင်ပါတယ်။ (Date variables တွေ ဖန်တီးဖို့အတွက် Data menu ပေါ်က Define Dates option ကို အသုံးပြုနိုင်ပါတယ်)။

Partial Least Squares Regression (တစ်စိတ်တစ်ပိုင်း အနည်းဆုံး နှစ်ထပ်ကိန်းများဖြင့် ဆက်စပ်မှု ခွဲခြမ်းစိတ်ဖြာခြင်း)

Partial Least Squares Regression (PLS) လုပ်ဆောင်ချက်က partial least squares regression မော်ဒယ်တွေကို ခန့်မှန်းတွက်ချက်ပေးပါတယ်။ (PLS ကို "projection to latent structure" လို့လည်း ခေါ်ပါတယ်)။ PLS ဟာ သာမန် ordinary least squares (OLS) regression, canonical correlation, ဒါမှမဟုတ် structural equation modeling တို့အစား အသုံးပြုနိုင်တဲ့ ကြိုတင်ခန့်မှန်းတွက်ချက်နိုင်တဲ့ နည်းလမ်း (predictive technique) တစ်ခုဖြစ်ပါတယ်။ အထူးသဖြင့် ခန့်မှန်းကိန်း (predictor variables) တွေဟာ အချင်းချင်း အရမ်းဆက်စပ်နေတဲ့ (highly correlated) အခါမျိုးမှာ ဒါမှမဟုတ် predictors အရေအတွက်က cases အရေအတွက်ထက် များနေတဲ့အခါမျိုးမှာ ဒီနည်းလမ်းက အရမ်းအသုံးဝင်ပါတယ်။

PLS ဟာ principal components analysis နဲ့ multiple regression တို့ရဲ့ လုပ်ဆောင်ချက်တွေကို ပေါင်းစပ်ထားတာပါ။ သူက independent နဲ့ dependent variables တွေကြားက ကွဲလွဲဆက်စပ်မှု (covariance) ကို တတ်နိုင်သမျှ အများဆုံး ရှင်းပြပေးနိုင်မယ့် မမြင်ရတဲ့ လျှို့ဝှက်အချက်အလက် (latent factors) အစုအဝေးတစ်ခုကို အရင်ဆုံး ထုတ်ယူပါတယ်။ အဲဒီနောက်မှာမှ independent variables တွေကို ဖြို့ခွဲထားတာကို သုံးပြီး dependent variables တွေရဲ့ တန်ဖိုးတွေကို ခန့်မှန်းတွက်ချက်ဖို့ regression အဆင့်ကို လုပ်ဆောင်တာ ဖြစ်ပါတယ်။

Tables (ဇယားများ)

ပုံမှန်အားဖြင့်တော့ (by default) ရှင်းပြနိုင်တဲ့ ကွဲလွဲမှုအချိုးအစား (latent factor အလိုက်)၊ latent factor အလေးချိန်များ (weights)၊ latent factor loadings တွေ၊ projection မှာ independent variable ရဲ့ အရေးပါမှု (VIP) နဲ့ (dependent variable အလိုက်) regression parameter ခန့်မှန်းချက်တွေ အားလုံးကို ထုတ်ပေးပါတယ်။

Charts (ပုံပြဇယားများ)။

Options tab ကနေတစ်ဆင့် projection မှာ variable ရဲ့ အရေးပါမှု (VIP)၊ factor ရမှတ်များ၊ ပထမဆုံး latent factors သုံးခုအတွက် factor အလေးချိန်များနဲ့ မော်ဒယ်ဆီသို့ အကွာအဝေး (distance to the model) တို့ကို ထုတ်ပေးပါတယ်။

Data considerations (Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ)

- **Measurement level (တိုင်းတာမှုအဆင့်)။** Dependent နဲ့ independent (predictor) variables တွေဟာ scale, nominal ဒါမှမဟုတ် ordinal ဖြစ်နိုင်ပါတယ်။ လုပ်ဆောင်ချက်က variables အားလုံးအတွက် သင့်လျော်တဲ့ measurement level ကို သတ်မှတ်ပေးပြီးသားလို့ ယူဆထားပါတယ်။ ဒါပေမယ့် မူရင်း variable စာရင်းထဲက variable ပေါ်မှာ right-click နှိပ်ပြီး ပေါ်လာတဲ့ menu ကနေ measurement level ကို ယာယီ ပြောင်းလဲလို့တော့

ရပါတယ်။ အမျိုးအစားခွဲ (categorical - nominal သို့မဟုတ် ordinal) variables တွေကိုတော့ ဒီလုပ်ဆောင်ချက်က တန်းတူညီမျှပဲ သဘောထားပြီး ကိုင်တွယ်သွားမှာပါ။

- **Categorical variable coding (အမျိုးအစားခွဲ ပြောင်းလဲကိန်းများကို ကုဒ်သတ်မှတ်ခြင်း)။** ဒီလုပ်ဆောင်ချက် အလုပ်လုပ်နေစဉ်အတွင်းမှာ categorical dependent variables တွေကို one-of-c coding သုံးပြီး ယာယီ recode (ကုဒ်ပြောင်း) လုပ်ပါတယ်။ တကယ်လို့ variable တစ်ခုမှာ category အရေအတွက် c ခု ရှိတယ်ဆိုရင်၊ အဲဒီ variable ကို c vectors အနေနဲ့ သိမ်းဆည်းပါတယ်။ ပထမဆုံး category ကို (1,0,...,0)၊ ဒုတိယတစ်ခုကို (0,1,0,...,0) နဲ့ နောက်ဆုံးတစ်ခုကို (0,0,...,0,1) ဆိုပြီး အသီးသီး ကိုယ်စားပြု သတ်မှတ်ပါတယ်။ Categorical dependent variables တွေကို dummy coding သုံးပြီး ကိုယ်စားပြုပါတယ် (ဆိုလိုတာက reference category နဲ့ သက်ဆိုင်တဲ့ indicator ကို ရိုးရိုးရှင်းရှင်း ဖယ်လှပ်ထားလိုက်တာပါ)။

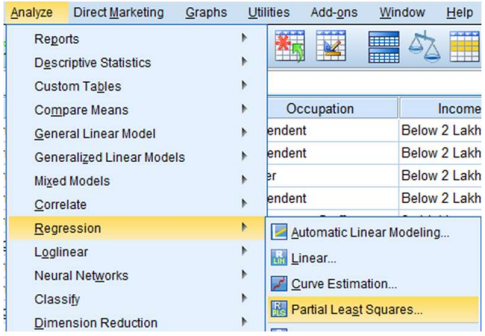
- **Frequency weights (အကြိမ်အရေအတွက် အလေးချိန်များ) ။** အလေးချိန် (Weight) တန်ဖိုးတွေကို အသုံးမပြုခင် အနီးစပ်ဆုံး ကိန်းပြည့်ဖြစ်အောင် (rounded) လုပ်ပါတယ်။ Weight တွေ missing ဖြစ်နေတဲ့ ဒါမှမဟုတ် weight တန်ဖိုး 0.5 ထက် ငယ်နေတဲ့ cases တွေကို analysis မှာ အသုံးမပြုပါဘူး။

- **Missing values (ပျောက်ဆုံးတန်ဖိုးများ)။** User-missing နဲ့ system-missing တန်ဖိုးတွေကို အကျုံးမဝင်ဘူး (invalid) လို့ သတ်မှတ်ပါတယ်။

- **Rescaling (စကေးပြန်လည်သတ်မှတ်ခြင်း)။** Categorical variables တွေကို ကိုယ်စားပြုတဲ့ indicator variables တွေ အပါအဝင် မော်ဒယ်ထဲက variables တွေ အားလုံးကို center နဲ့ standardized အဖြစ် ပြောင်းလဲပေးပါတယ်။

Obtaining Partial Least Squares Regression (Partial Least Squares Regression ရယူအသုံးပြုရန် အဆင့်များ)

Menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Regression > Partial Least Squares...



၁။ အနည်းဆုံး dependent variable တစ်ခုကို ရွေးချယ်ပါ။

၂။ အနည်းဆုံး independent variable တစ်ခုကို ရွေးချယ်ပါ။

ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့ သင် အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- Categorical (nominal ဒါမှမဟုတ် ordinal) dependent variables တွေအတွက် reference category တစ်ခုကို သတ်မှတ်နိုင်ပါတယ်။
- Casewise output တွေနဲ့ သိမ်းဆည်းထားတဲ့ datasets တွေအတွက် သီးခြားခွဲခြားသိနိုင်မယ့် အမှတ်အသား (unique identifier) အဖြစ် သုံးဖို့ variable တစ်ခုကို သတ်မှတ်နိုင်ပါတယ်။
- ထုတ်ယူမယ့် latent factors အရေအတွက်အတွက် အမြင့်ဆုံးကန့်သတ်ချက် (upper limit) ကို သတ်မှတ်နိုင်ပါတယ်။

Prerequisites (ကြိုတင်လိုအပ်ချက်များ)

Partial Least Squares Regression လုပ်ဆောင်ချက်ဟာ Python extension command တစ်ခုဖြစ်ပြီး၊ သင်ရဲ့ IBM SPSS Statistics ထုတ်ကုန်နဲ့အတူ default အနေနဲ့ install လုပ်ပြီးသားဖြစ်တဲ့ IBM SPSS Statistics - Essentials for Python ကို လိုအပ်ပါတယ်။ ဒါ့အပြင် အခမဲ့ရယူနိုင်တဲ့ NumPy နဲ့ SciPy Python libraries တွေကိုလည်း လိုအပ်ပါတယ်။

(မှတ်ချက် - Distributed analysis mode နဲ့ အလုပ်လုပ်နေတဲ့ user တွေအတွက်တော့ NumPy နဲ့ SciPy ကို server ပေါ်မှာ install လုပ်ထားရပါမယ်။ Windows နဲ့ Mac သုံးစွဲသူတွေကတော့ Python 3.9 version သီးသန့်တစ်ခုမှာ ၎င်းတို့ကို install လုပ်ပြီး Options dialog ထဲက File Locations မှာ လမ်းကြောင်း ညွှန်ပေးရမှာ ဖြစ်ပါတယ်။ Linux နဲ့ Unix Server သုံးစွဲသူတွေအတွက်လည်း အလားတူ Python version 3.9 ပေါ်မှာ သီးခြား install လုပ်ဖို့ လိုအပ်ပါတယ်။)

Model (မော်ဒယ်)

- **Specify Model Effects (မော်ဒယ်သက်ရောက်မှုများ သတ်မှတ်ခြင်း)** ။ Main-effects မော်ဒယ်တစ်ခုမှာ factor နဲ့ covariate တွေရဲ့ အဓိကသက်ရောက်မှု (main effects) အားလုံးပါဝင်ပါတယ်။ အချင်းချင်းသက်ရောက်မှု (interactions) တွေကို သတ်မှတ်ဖို့ Custom ကို ရွေးချယ်ပါ။ မော်ဒယ်ထဲ ထည့်သွင်းချင်တဲ့ terms အားလုံးကို သင်ကိုယ်တိုင် ညွှန်ပြပေးရပါမယ်။
- **Factors and Covariates**။ Factors နဲ့ Covariates တွေကို စာရင်းပြထားပါတယ်။

- **Model။** Model ကတော့ သင့် data ရဲ့ သဘာဝအပေါ် မူတည်ပါတယ်။ Custom ကို ရွေးပြီးတဲ့နောက်မှာ သင့် analysis အတွက် စိတ်ဝင်စားစရာကောင်းတဲ့ main effects နဲ့ interactions တွေကို သင်ရွေးချယ်နိုင်ပါတယ်။

Build Terms (သတ်မှတ်ချက်များ တည်ဆောက်ခြင်း) ရွေးချယ်ထားတဲ့ factors နဲ့ covariates တွေအတွက် အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- **Interaction (အချင်းချင်း သက်ရောက်မှု) ။** ရွေးချယ်ထားတဲ့ variables အားလုံးရဲ့ အမြင့်ဆုံးအဆင့် (highest-level) interaction term ကို ဖန်တီးပေးပါတယ်။ ဒါက ပုံမှန် (default) အပြင်အဆင်ပါ။

- **Main effects (အဓိက သက်ရောက်မှုများ) ။** ရွေးချယ်ထားတဲ့ variable တစ်ခုချင်းစီအတွက် main-effects term တစ်ခုစီ ဖန်တီးပေးပါတယ်။

- **All 2-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၂-ခုတွဲ (two-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။

- **All 3-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၃-ခုတွဲ (three-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။

- **All 4-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၄-ခုတွဲ (four-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။

- **All 5-way။** ရွေးချယ်ထားတဲ့ variables တွေရဲ့ ဖြစ်နိုင်သမျှ ၅-ခုတွဲ (five-way) interactions အားလုံးကို ဖန်တီးပေးပါတယ်။

Options (ရွေးချယ်စရာများ)

Options tab က အသုံးပြုသူတွေကို individual cases, latent factors, နဲ့ predictors တွေအတွက် မော်ဒယ်ခန့်မှန်းချက်တွေကို သိမ်းဆည်းဖို့ (save) နဲ့ ပုံဆွဲပြဖို့ (plot) ခွင့်ပြုပေးပါတယ်။ Data အမျိုးအစား တစ်ခုစီအတွက် dataset ရဲ့ နာမည်ကို သတ်မှတ်ပေးပါ။ Dataset နာမည်တွေက တစ်ခုနဲ့တစ်ခု မတူဘဲ ထူးခြား (unique) နေရပါမယ်။ တကယ်လို့ ရှိပြီးသား dataset နာမည်ကို ပေးလိုက်ရင် သူ့ရဲ့ မူလအချက်အလက်တွေ နေရာမှာ အစားထိုး ဝင်ရောက်သွားမှာဖြစ်ပြီး၊ မရှိသေးတဲ့ နာမည်ပေးရင်တော့ dataset အသစ်တစ်ခု ဖန်တီးပေးမှာ ဖြစ်ပါတယ်။

- **Save estimates for individual cases (case တစ်ခုချင်းစီအတွက် ခန့်မှန်းချက်များကို သိမ်းဆည်းခြင်း) ။** ၎င်းက predicted values (ခန့်မှန်းတန်ဖိုးများ)၊ residuals (ကွာဟချက်များ)၊ latent factor မော်ဒယ်ဆီသို့ အကွာအဝေး၊ နဲ့ latent factor ရမှတ်များ

စတဲ့ casewise မော်ဒယ်ခန့်မှန်းချက်တွေကို သိမ်းဆည်းပေးပါတယ်။ Latent factor ရုမှတ်တွေကိုလည်း ပုံဆွဲ (plot) ပေးပါသေးတယ်။

- **Save estimates for latent factors (latent factors များအတွက် ခန့်မှန်းချက်များကို သိမ်းဆည်းခြင်း)** ။ ၎င်းက latent factor loadings တွေနဲ့ latent factor weights (အလေးချိန်များ) ကို သိမ်းဆည်းပေးပါတယ်။ Latent factor weights တွေကိုလည်း ပုံဆွဲပေးပါတယ်။
- **Save estimates for independent variables (independent variables များအတွက် ခန့်မှန်းချက်များကို သိမ်းဆည်းခြင်း)** ။ ၎င်းက regression parameter ခန့်မှန်းချက်တွေနဲ့ projection မှာ variable ရဲ့ အရေးပါမှု (VIP) ကို သိမ်းဆည်းပေးပါတယ်။ VIP ကို latent factor အလိုက် ပုံဆွဲပေးပါသေးတယ်။

Ratio Statistics (အချိုးအစားဆိုင်ရာ စာရင်းအင်းများ)

Ratio Statistics (အချိုးအစားဆိုင်ရာ စာရင်းအင်းများ) လုပ်ဆောင်ချက်က scale variables (စကေးပြောင်းလဲကိန်း) နှစ်ခုကြားက အချိုးအစားကို ဖော်ပြပေးဖို့အတွက် ပြီးပြည့်စုံတဲ့ အကျဉ်းချုပ် စာရင်းအင်း (summary statistics) စာရင်းကို ထောက်ပံ့ပေးပါတယ်။ သင်က ထွက်လာတဲ့ ရလဒ် (output) ကို grouping variable (အုပ်စုခွဲပြောင်းလဲကိန်း) ရဲ့ တန်ဖိုးတွေအပေါ် မူတည်ပြီး အငယ်မှအကြီး ဒါမှမဟုတ် အကြီးမှအငယ် အစဉ်လိုက် စီစဉ် (sort) နိုင်ပါတယ်။ Ratio statistics အစီရင်ခံစာကို output ထဲမှာ ဖျောက်ထား (suppress) နိုင်သလို၊ ရလဒ်တွေကို ပြင်ပဖိုင် (external file) တစ်ခုအနေနဲ့လည်း သိမ်းဆည်းနိုင်ပါတယ်။

Example (ဥပမာ)

ခရိုင်ငါးခုစီမှာရှိတဲ့ အိမ်တွေရဲ့ ခန့်မှန်းတန်ဖိုး (appraisal price) နဲ့ ရောင်းဈေး (sale price) ကြားက အချိုးအစားမှာ ကောင်းမွန်တဲ့ တူညီညီမျှမှု (uniformity) ရှိပါသလား။ ရရှိလာတဲ့ output ကနေတစ်ဆင့် အချိုးအစားတွေရဲ့ ဖြန့်ကျက်မှု (distribution of ratios) ဟာ ခရိုင်တစ်ခုနဲ့တစ်ခု သိသိသာသာ ကွဲပြားခြားနားနေတယ်ဆိုတာကို သင် သိရှိနိုင်ပါတယ်။

Statistics (စာရင်းအင်းများ)

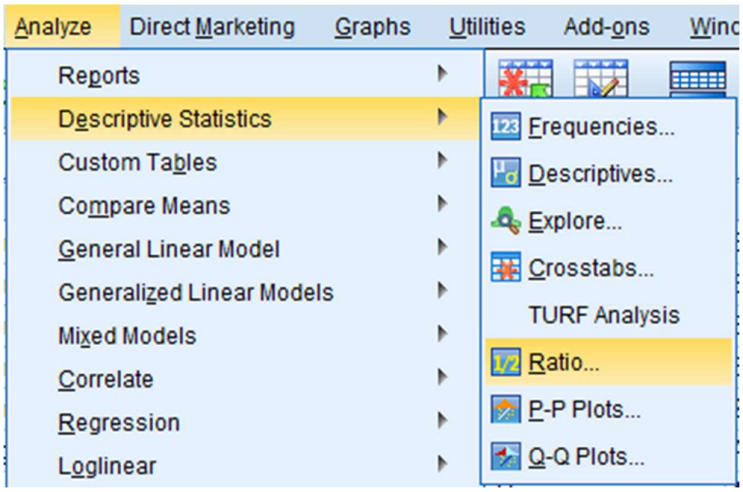
Median, mean, weighted mean, confidence intervals, coefficient of dispersion (COD), coefficient of variation (COV), median-centered coefficient of variation, mean-centered coefficient of variation, price-related bias (PRB), price-related differential (PRD), standard deviation, average absolute deviation (AAD), range, minimum နဲ့ maximum တန်ဖိုးတွေအပြင်၊ အသုံးပြုသူသတ်မှတ်ထားတဲ့ range ဒါမှမဟုတ် median ratio အတွင်းက ရာခိုင်နှုန်းတစ်ခုအတွက် တွက်ချက်ထားတဲ့ concentration index တို့ ပါဝင်ပါတယ်။

Ratio Statistics Data Considerations (Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ)

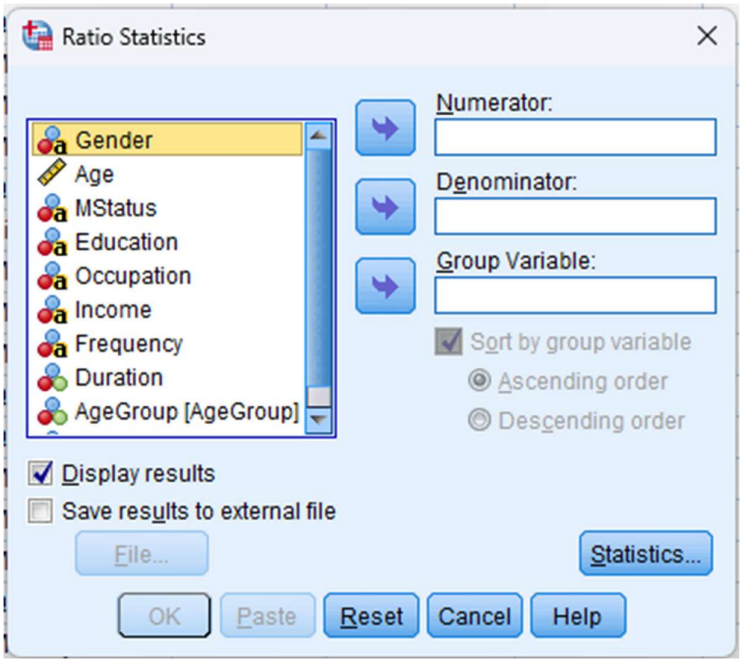
- **Data (အချက်အလက်)။** Grouping variables (nominal သို့မဟုတ် ordinal အဆင့် တိုင်းတာမှုများ) တွေကို ကုဒ်သတ်မှတ်ဖို့ (code လုပ်ဖို့) ဂဏန်းကုဒ်တွေ (numeric codes) ဒါမှမဟုတ် စာသားတွေ (strings) ကို အသုံးပြုပါ။
- **Assumptions (ယူဆချက်များ)။** အချိုးတစ်ခုရဲ့ ပိုင်းဝေ (numerator) နဲ့ ပိုင်းခြေ (denominator) ကို သတ်မှတ်ပေးတဲ့ variables တွေဟာ အပေါင်းလက္ခဏာတန်ဖိုး (positive values) တွေရှိတဲ့ scale variables တွေ ဖြစ်သင့်ပါတယ်။

Obtaining Ratio Statistics (Ratio Statistics ရယူအသုံးပြုရန် အဆင့်များ)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Descriptive Statistics > Ratio...



၂။ ပိုင်းဝေ (numerator) variable တစ်ခုကို ရွေးချယ်ပါ။ ၃။ ပိုင်းခြေ (denominator) variable တစ်ခုကို ရွေးချယ်ပါ။



ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့ အောက်ပါတို့ကို လုပ်ဆောင်နိုင်ပါတယ်-

- Grouping variable တစ်ခုကို ရွေးချယ်ပြီး ရလဒ်တွေထဲမှာ အုပ်စုတွေရဲ့ အစီအစဉ် (ordering) ကို သတ်မှတ်နိုင်ပါတယ်။

- ရလဒ်တွေကို Viewer ထဲမှာ ပြသမလား ဆိုတာကို ရွေးချယ်နိုင်ပါတယ်။
- PRINT subcommand မှာ N keyword တစ်ခု ထည့်သွင်းဖို့ ရွေးချယ်နိုင်ပါတယ်။ ဒီအပြင်အဆင်က output ထဲက statistics table ထဲကို sample sizes တွေ ပေါင်းထည့်ပေးပါတယ်။
- ရလဒ်တွေကို နောက်ပိုင်းပြန်သုံးဖို့ ပြင်ပဖိုင် (external file) တစ်ခုအနေနဲ့ သိမ်းဆည်းမလားဆိုတာကို ရွေးချယ်နိုင်ပြီး၊ ရလဒ်တွေ သိမ်းဆည်းမယ့် ဖိုင်နာမည်ကို သတ်မှတ်နိုင်ပါတယ်။

Ratio Statistics: Statistics (Ratio Statistics ဆိုင်ရာ စာရင်းအင်းများ)

Central Tendency (ဗဟိုသို့တိမ်းညွတ်မှု) ဗဟိုသို့တိမ်းညွတ်မှု အတိုင်းအတာတွေဟာ အချိုးတွေရဲ့ ဖြန့်ကျက်မှု (distribution of ratios) ကို ဖော်ပြပေးတဲ့ စာရင်းအင်းတွေ ဖြစ်ပါတယ်။

- **Median (အလယ်ကိန်း)။** ဒီတန်ဖိုးထက် ငယ်တဲ့ အချိုးအရေအတွက်နဲ့ ဒီတန်ဖိုးထက် ကြီးတဲ့ အချိုးအရေအတွက် တူညီနေစေမယ့် တန်ဖိုးတစ်ခု ဖြစ်ပါတယ်။
- **Mean (ပျမ်းမျှကိန်း)။** အချိုးတွေအားလုံးကို ပေါင်းပြီး ရလာတဲ့ရလဒ်ကို စုစုပေါင်း အချိုးအရေအတွက်နဲ့ စားထားတဲ့ ရလဒ်ဖြစ်ပါတယ်။
- **Weighted Mean (အလေးချိန်ပေး ပျမ်းမျှကိန်း) ။** ပိုင်းဝေရဲ့ mean ကို ပိုင်းခြေရဲ့ mean နဲ့ စားထားတဲ့ ရလဒ်ဖြစ်ပါတယ်။ Weighted mean ဟာ ပိုင်းခြေနဲ့ အလေးချိန်ပေးထားတဲ့ (weighted by the denominator) အချိုးတွေရဲ့ mean လည်း ဖြစ်ပါတယ်။

Confidence intervals for PRB and measures of central tendency (PRB နှင့် ဗဟိုတိမ်းညွတ်မှု အတိုင်းအတာများအတွက် ယုံကြည်မှု ကြားကာလများ)

PRB, mean, median နဲ့ (တောင်းဆိုထားခဲ့ရင်) weighted mean တို့အတွက် confidence intervals တွေကို ပြသပေးပါတယ်။ Confidence level အဖြစ် 0 နဲ့ ညီတာ ဒါမှမဟုတ် 0 ထက်ကြီးပြီး 100 ထက်ငယ်တဲ့ တန်ဖိုးတစ်ခုကို သတ်မှတ်ပေးပါ။ ပုံမှန် (default) အပြင်အဆင်ကတော့ 95% ဖြစ်ပါတယ်။

Dispersion (ပြန့်ကျဲမှု)

ဒီစာရင်းအင်းတွေက လေ့လာတွေ့ရှိရတဲ့ တန်ဖိုးတွေထဲက ပြောင်းလဲမှု ဒါမှမဟုတ် ပြန့်ကျဲမှု ပမာဏ (amount of variation, or spread) ကို တိုင်းတာပေးပါတယ်။

- **AAD။** Average absolute deviation ဆိုတာ အချိုးတွေရဲ့ median အပေါ် မူတည်ပြီး ကွဲလွဲနေတဲ့ absolute deviations တွေကို ပေါင်းကာ၊ ရလာတဲ့ရလဒ်ကို စုစုပေါင်း အချိုးအရေအတွက်နဲ့ စားထားတဲ့ ရလဒ် ဖြစ်ပါတယ်။

- **COD။** Coefficient of dispersion ဆိုတာ average absolute deviation ကို median ရဲ့ ရာခိုင်နှုန်းတစ်ခုအနေနဲ့ ဖော်ပြထားတဲ့ ရလဒ် ဖြစ်ပါတယ်။

- **COV။** Coefficient of variation ဖြစ်ပါတယ်။ Median-centered coefficient of variation ဆိုတာ median ကနေ ကွဲလွဲမှုတွေရဲ့ root mean squares ကို median ရဲ့ ရာခိုင်နှုန်းတစ်ခုအနေနဲ့ ဖော်ပြထားတဲ့ ရလဒ်ပါ။ Mean-centered coefficient of variation ကတော့ standard deviation ကို mean ရဲ့ ရာခိုင်နှုန်းတစ်ခုအနေနဲ့ ဖော်ပြထားတဲ့ ရလဒ် ဖြစ်ပါတယ်။

- **PRB။** Price-related bias (ဈေးနှုန်းနှင့်ဆက်စပ်သော ဘက်လိုက်မှု) ဖြစ်ပါတယ်။ တန်ဖိုးပိုကြီးတဲ့ ပိုင်ဆိုင်မှုတွေအတွက် အကဲဖြတ်တန်ဖိုးနဲ့ ဈေးနှုန်း အချိုးအစား (assessment to price ratios) တွေဟာ စနစ်တကျ ပိုများနေသလား သို့မဟုတ် ပိုနည်းနေသလား ဆိုတာကို ပြတဲ့ အညွှန်းကိန်းတစ်ခုပါ။ အရောင်းဈေးနှုန်းတွေရဲ့ ပျမ်းမျှနဲ့ အကဲဖြတ်တန်ဖိုးတွေကို median အချိုးနဲ့စားထားတဲ့ အချိုးတွေအဖြစ် တွက်ချက်ထားတဲ့ value proxy measures တွေရဲ့ base 2 logarithms အပေါ်မှာ မူတည်ပြီး၊ assessment ratios တွေရဲ့ ရာခိုင်နှုန်း ကွာခြားချက်တွေကို regress လုပ်ပါတယ်။ တန်ဖိုး 100 ရာခိုင်နှုန်း ပြောင်းလဲသွားတဲ့အခါ assessment ratios တွေမှာ ဖြစ်ပေါ်လာမယ့် ရာခိုင်နှုန်း အပြောင်းအလဲကို ဖော်ပြပေးပါတယ်။

- **PRD။** Price-related differential ဖြစ်ပြီး သူ့ကို index of regressivity လို့လည်း ခေါ်ပါတယ်၊ mean ကို weighted mean နဲ့ စားထားတဲ့ ရလဒ် ဖြစ်ပါတယ်။

- **Standard deviation (စံသွေဖည်ကိန်း)။** Standard deviation ဟာ အချိုးတွေရဲ့ mean ကနေ သွေဖည်မှု နှစ်ထပ်ကိန်း (squared deviations) တွေကို ပေါင်းပြီး၊ ရလာတဲ့ရလဒ်ကို စုစုပေါင်းအချိုးအရေအတွက်ထဲကနေ တစ်နှုတ်ထားတဲ့ကိန်း (total number of ratios minus one) နဲ့စားကာ၊ အပေါင်းလက္ခဏာရှိတဲ့ square root (နှစ်ထပ်ကိန်းရင်း) ကို ယူထားတဲ့ ရလဒ် ဖြစ်ပါတယ်။

- **Range (တန်ဖိုးပြန့်ကျက်မှု အတိုင်းအတာ)။** Range ဟာ အကြီးဆုံးအချိုး (maximum ratio) ထဲကနေ အငယ်ဆုံးအချိုး (minimum ratio) ကို နှုတ်ထားတဲ့ ရလဒ် ဖြစ်ပါတယ်။

- **Minimum။** အငယ်ဆုံး အချိုး ဖြစ်ပါတယ်။

- **Maximum။** အကြီးဆုံး အချိုး ဖြစ်ပါတယ်။

Concentration Index (စုစည်းမှု အညွှန်းကိန်း)

Coefficient of concentration က သတ်မှတ်ထားတဲ့ ကြားကာလ (interval) တစ်ခုအတွင်းမှာ ကျရောက်နေတဲ့ အချိုးတွေရဲ့ ရာခိုင်နှုန်းကို တိုင်းတာပေးပါတယ်။ သူ့ကို မတူညီတဲ့ နည်းလမ်းနှစ်မျိုးနဲ့ တွက်ချက်နိုင်ပါတယ်-

- **Between Proportions (အချိုးအစားများအကြား)။** ဒီနေရာမှာတော့ interval ကို သူ့ရဲ့ အနိမ့်ဆုံးနဲ့ အမြင့်ဆုံး တန်ဖိုးတွေကို သတ်မှတ်ပေးခြင်းအားဖြင့် တိတိကျကျ သတ်မှတ်ပါတယ်။ Low proportion နဲ့ high proportion တွေအတွက် တန်ဖိုးတွေကို ရိုက်ထည့်ပြီး၊ interval တစ်ခုရရှိဖို့ Add ကို နှိပ်ပါ။
- **Within Percentage of Median (Median ၏ ရာခိုင်နှုန်းအတွင်း) ။** ဒီနေရာမှာတော့ interval ကို median ရဲ့ ရာခိုင်နှုန်းကို သတ်မှတ်ပေးခြင်းအားဖြင့် သွယ်ဝိုက်ပြီး သတ်မှတ်ပါတယ်။ 0 နဲ့ 100 ကြား တန်ဖိုးတစ်ခုကို ရိုက်ထည့်ပြီး Add ကို နှိပ်ပါ။ Interval ရဲ့ အောက်ခြေအဆုံးသတ် (lower end) ဟာ (1 – 0.01 × value) × median နဲ့ ညီမျှပြီး၊ အပေါ်ဘက်အဆုံးသတ် (upper end) ကတော့ (1 + 0.01 × value) × median နဲ့ ညီမျှပါတယ်။

P-P Plot (P-P ဖြစ်တန်စွမ်း ပုံပြဇယားများ)

P-P plots လုပ်ဆောင်ချက်ဟာ sequence (အစီအစဉ်) ဒါမှမဟုတ် time series (အချိန်စဉ်ဆက်) ပြောင်းလဲကိန်း (variables) တစ်ခု သို့မဟုတ် တစ်ခုထက်ပိုတဲ့ ဟာတွေအတွက် probability plots (ဖြစ်တန်စွမ်း ပုံပြဇယားများ) ကို ထုတ်ပေးပါတယ်။ အဲဒီ variables တွေကို ဇယားမဆွဲခင်မှာ standardized (စံသတ်မှတ်ခြင်း)၊ differenced (ကွာခြားချက်ရှာခြင်း) နဲ့ transformed (ပုံစံပြောင်းလဲခြင်း) တွေ ပြုလုပ်နိုင်ပါတယ်။

ရရှိနိုင်တဲ့ test distributions (စမ်းသပ်ဖြန့်ကျက်မှုပုံစံများ) ထဲမှာ beta, chi-square, exponential, gamma, half-normal, Laplace, Logistic, Lognormal, normal, pareto, Student's t, Weibull, နဲ့ uniform တို့ ပါဝင်ပါတယ်။ သင်ရွေးချယ်လိုက်တဲ့ distribution ပေါ်မူတည်ပြီး degrees of freedom နဲ့ အခြား parameters တွေကို သတ်မှတ်ပေးနိုင်ပါတယ်။

- **Transformed values (ပုံစံပြောင်းထားတဲ့ တန်ဖိုးများ)** အတွက်လည်း probability plots တွေကို သင်ရယူနိုင်ပါတယ်။ Transformation options (ပုံစံပြောင်းလဲမှုဆိုင်ရာ ရွေးချယ်စရာများ) အနေနဲ့ natural log, standardize values, difference, နဲ့ seasonally difference တို့ ပါဝင်ပါတယ်။

- **Expected distributions (မျှော်မှန်းထားတဲ့ ဖြန့်ကျက်မှုများ)** ကို တွက်ချက်မယ့် နည်းလမ်းနဲ့ "ties" (တန်ဖိုးတူညီနေတဲ့ လေ့လာတွေ့ရှိချက်များ) ကို ဖြေရှင်းမယ့် နည်းလမ်းတွေကိုလည်း သင်သတ်မှတ်ပေးနိုင်ပါတယ်။

Test distribution (စမ်းသပ်မည့် ဖြန့်ကျက်မှုပုံစံ)

သင့် data အတွက် distribution အမျိုးအစားကို သတ်မှတ်ပေးပါ။ Drop-down list မှာ အောက်ပါတို့ကို ရွေးချယ်နိုင်ပါတယ်-

- **Beta:** Beta distribution ဖြစ်ပါတယ်။ shape1 နဲ့ shape2 parameters တွေဖြစ်တဲ့ a နဲ့ b က အပေါင်းလက္ခဏာ (positive) ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် DISTRIBUTION ကနေ သူတို့ကို sample mean နဲ့ sample standard deviation ကိုသုံးပြီး ခန့်မှန်းပေးပါတယ်။ Observations အားလုံးက 0 နဲ့ 1 ကြားမှာပဲ ရှိရပါမယ်။

- **Chi-square:** Chi-square distribution ဖြစ်ပါတယ်။ Degrees of freedom (df) ကို သတ်မှတ်ပေးရပါမယ်။ အနှုတ်လက္ခဏာပါတဲ့ observations တွေကို လက်မခံပါဘူး။

- **Exponential:** Exponential distribution ဖြစ်ပါတယ်။ Scale parameter a ဟာ အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် sample mean ကနေ ခန့်မှန်းပေးပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။

- **Gamma:** Gamma distribution ဖြစ်ပါတယ်။ Shape နဲ့ scale parameters (a နဲ့ b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် sample mean နဲ့ sample standard deviation ကနေ ခန့်မှန်းပေးပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။
- **Half-normal:** Half-normal distribution ဖြစ်ပါတယ်။ Data တွေကို location-free သို့မဟုတ် centralized ဖြစ်တယ်လို့ ယူဆထားပါတယ်။ Scale parameter a ကို သတ်မှတ်နိုင်သလို maximum likelihood method နဲ့လည်း ခန့်မှန်းခိုင်းလို့ရပါတယ်။
- **Laplace:** Laplace သို့မဟုတ် double exponential distribution ဖြစ်ပါတယ်။ Location နဲ့ scale parameter (a နဲ့ b) လိုအပ်ပြီး b က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် sample mean နဲ့ sample standard deviation ကနေ ခန့်မှန်းပေးပါတယ်။
- **Logistic:** Logistic distribution ဖြစ်ပါတယ်။ Location နဲ့ scale parameter (a နဲ့ b) ပါဝင်ပြီး b က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် sample mean နဲ့ sample standard deviation ကနေ ခန့်မှန်းပေးပါတယ်။
- **Lognormal:** Lognormal distribution ဖြစ်ပါတယ်။ Scale နဲ့ shape parameters (a နဲ့ b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် sample data ရဲ့ natural logarithm ရဲ့ mean နဲ့ standard deviation ကနေ ခန့်မှန်းပေးပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။
- **Normal:** Normal (ပုံမှန်) distribution ဖြစ်ပါတယ်။ Location parameter a ဟာ ဂဏန်းတန်ဖိုးတစ်ခုခုဖြစ်နိုင်ပြီး scale parameter b ကတော့ အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် sample mean နဲ့ sample standard deviation ကနေ ခန့်မှန်းပေးပါတယ်။
- **Pareto:** Pareto distribution ဖြစ်ပါတယ်။ Threshold နဲ့ shape parameters (a နဲ့ b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် a ကို အငယ်ဆုံး observation နဲ့ ညီတယ်လို့ ယူဆပြီး b ကိုတော့ maximum likelihood method နဲ့ ခန့်မှန်းပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။
- **Student t:** Student's t distribution ဖြစ်ပါတယ်။ Degrees of freedom (df) ကို သတ်မှတ်ပေးရပါမယ်။
- **Uniform:** Uniform distribution ဖြစ်ပါတယ်။ Minimum နဲ့ maximum parameter (a နဲ့ b) ပါဝင်ပြီး a က b ထက် ကြီးတာ ဒါမှမဟုတ် ညီနေရပါမယ်။ မသတ်မှတ်ထားရင် sample data ကနေ ယူဆတွက်ချက်ပေးပါတယ်။

- **Weibull:** Weibull distribution ဖြစ်ပါတယ်။ Scale နဲ့ shape parameters (a နဲ့ b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် least square method နဲ့ ခန့်မှန်းပေးပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။

Distribution parameters (ဖြန့်ကျက်မှုဆိုင်ရာ ပါရာမီတာများ)

- **Estimate from data:** ဒါကိုရွေးထားရင် data နဲ့ ရွေးချယ်ထားတဲ့ distribution အမျိုးအစားပေါ် မူတည်ပြီး distribution parameters တွေကို ခန့်မှန်းပေးပါတယ်။
- **Specify:** ဒါကိုရွေးထားရင် ရွေးချယ်ထားတဲ့ distribution အတွက် parameters တွေကို သင်ကိုယ်တိုင် သတ်မှတ်ပေးနိုင်ပါတယ်။

Transform (ပုံစံပြောင်းလဲခြင်း)

- **Natural log transform:** Data ထဲက ပြောင်းလဲနေတဲ့ amplitude (လွှဲကျယ်) တွေကို ဖယ်ရှားဖို့ သဘာဝ လော့ဂရစ်သမ် (natural logarithm - base e) ကိုသုံးပြီး data ကို ပုံစံပြောင်းပေးပါတယ်။
- **Standardize values:** Sequence သို့မဟုတ် time series variables တွေကို mean 0 နဲ့ standard deviation 1 ရှိတဲ့ sample တစ်ခုအနေနဲ့ ပြောင်းလဲပေးပါတယ်။
- **Difference:** မတည်ငြိမ်တဲ့ (non-stationary) variable တစ်ခုကို ကိန်းသေ mean နဲ့ variance ရှိတဲ့ တည်ငြိမ်တဲ့ (stationary) variable အဖြစ် ပြောင်းလဲဖို့ ဇယားမဆွဲခင် အသုံးပြုမယ့် differencing အတိုင်းအတာကို သတ်မှတ်ပေးပါတယ်။
- **Seasonally difference:** Variable က ရာသီအလိုက် သို့မဟုတ် အချိန်အပိုင်းအခြားအလိုက် ပုံစံ (seasonal/periodic pattern) ပြနေရင်၊ ဇယားမဆွဲခင် seasonally difference လုပ်ဖို့ သုံးနိုင်ပါတယ်။

Proportion estimation formula (အချိုးအစား ခန့်မှန်းခြင်း ဖော်မြူလာ)

- **Blom's:** (r-3/8) / (w+1/4) ဖော်မြူလာကို သုံးပါတယ်၊ w က case weights စုစုပေါင်းဖြစ်ပြီး r က rank (အဆင့်) ဖြစ်ပါတယ်။
- **Rankit:** (r-1/2) / w ဖော်မြူလာကို သုံးပါတယ်။
- **Tukey's:** (r-1/3) / (w+1/3) ဖော်မြူလာကို သုံးပါတယ်။
- **Van der Waerden's:** r/(w+1) ဖော်မြူလာကို သုံးပါတယ်။

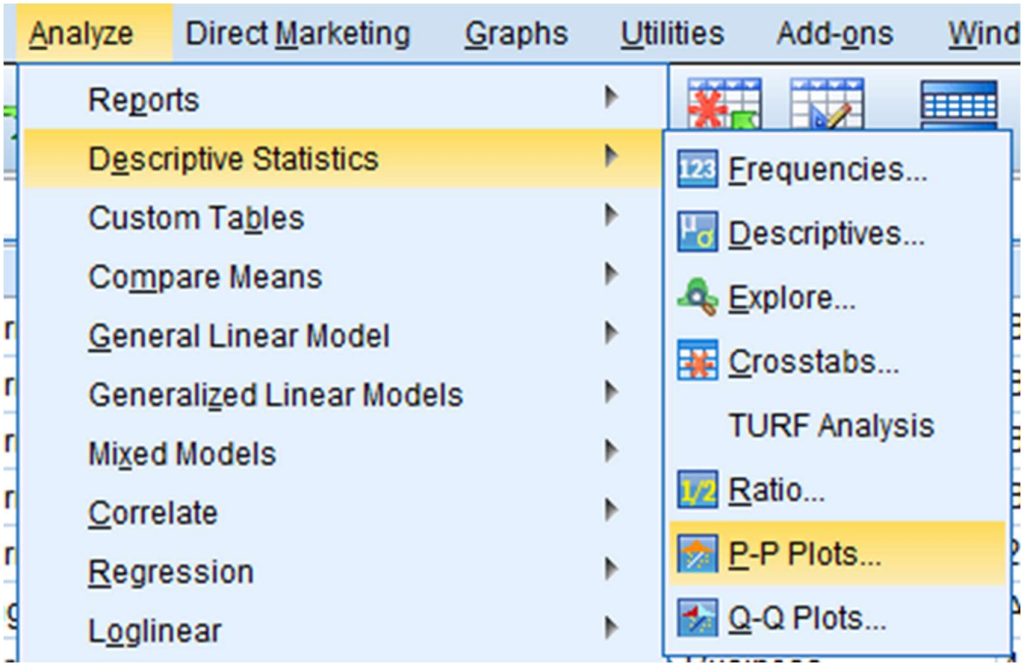
Rank assigned to ties (တန်ဖိုးတူညီမှုများအား အဆင့်သတ်မှတ်ခြင်း)

တန်ဖိုးတူညီနေတဲ့ (tied values) တွေကို အဆင့်သတ်မှတ်ရာမှာ ဘယ်လိုကိုင်တွယ်မလဲဆိုတာကို ထိန်းချုပ်ပေးပါတယ် (ဥပမာ- ဇယားအရ Mean, Low, High, Break ties arbitrarily စသဖြင့် သတ်မှတ်နိုင်ပါတယ်)။

Value	Mean	Low	High	Break ties arbitrarily
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

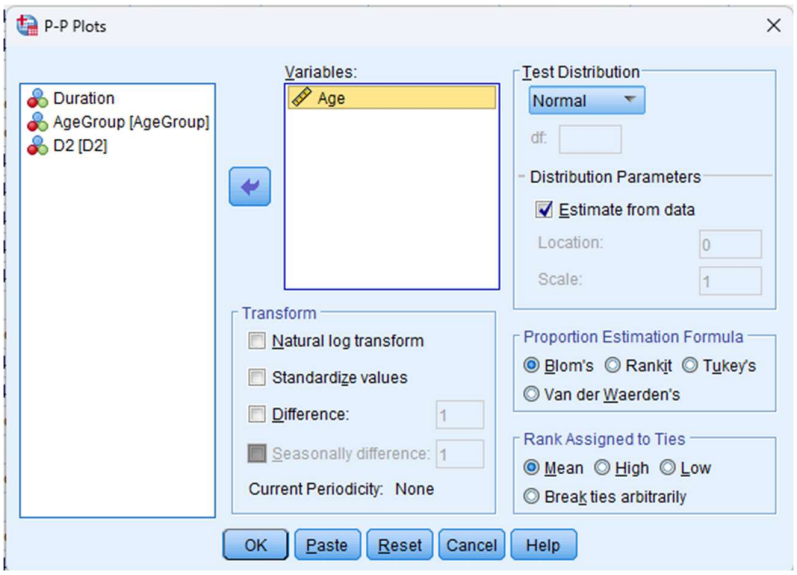
Obtaining P-P probability plots (P-P probability plots ရယူရန် အဆင့်များ)

၁။ Menus တွေထဲကနေ Analyze > Descriptive Statistics > P-P Plots... ကို ရွေးချယ်ပါ။



၂။ Numeric variables တစ်ခု သို့မဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ကာ Variables field ထဲ ထည့်ပါ။

၃။ Test distribution တစ်ခုကို ရွေးချယ်ပါ။ (လိုအပ်ပါက transformation options တွေနဲ့ မျှော်မှန်းထားတဲ့ ဖြန့်ကျက်မှု တွက်ချက်မယ့် နည်းလမ်းတွေကို ထပ်မံရွေးချယ်နိုင်ပါတယ်)။



Q-Q Plot (Q-Q ဖြစ်တန်စွမ်း ပုံပြဇယားများ)

Q-Q plots လုပ်ဆောင်ချက်ဟာ ပုံစံပြောင်းထားတဲ့ တန်ဖိုးတွေ (transformed values) အတွက် probability plots တွေကို ထုတ်ပေးပါတယ်။ ရရှိနိုင်တဲ့ စမ်းသပ်ဖြန့်ကျက်မှု (test distributions) ပုံစံတွေထဲမှာ beta, chi-square, exponential, gamma, half-normal, Laplace, Logistic, Lognormal, normal, pareto, Student's t, Weibull နဲ့ uniform တို့ ပါဝင်ပါတယ်။ သင်ရွေးချယ်လိုက်တဲ့ distribution အပေါ်မူတည်ပြီး degrees of freedom နဲ့ အခြား parameters တွေကို သတ်မှတ်ပေးနိုင်ပါတယ်။

- သင်က ပုံစံပြောင်းထားတဲ့ တန်ဖိုးတွေအတွက် probability plots တွေကို ရယူနိုင်ပါတယ်။ ပုံစံပြောင်းလဲမှုဆိုင်ရာ ရွေးချယ်စရာ (Transformation options) တွေအနေနဲ့ natural log, standardize values, difference နဲ့ seasonally difference တို့ ပါဝင်ပါတယ်။
- မျှော်မှန်းထားတဲ့ ဖြန့်ကျက်မှု (expected distributions) တွေကို တွက်ချက်မယ့် နည်းလမ်းနဲ့ "ties" (တန်ဖိုးတူညီနေတဲ့ လေ့လာတွေ့ရှိချက် အများအပြား) တွေကို ဖြေရှင်းမယ့် နည်းလမ်းတွေကိုလည်း သင်သတ်မှတ်ပေးနိုင်ပါတယ်။

Test distribution (စမ်းသပ်မည့် ဖြန့်ကျက်မှုပုံစံ)

သင့် data အတွက် distribution အမျိုးအစားကို သတ်မှတ်ပေးပါ။ Drop-down list မှာ အောက်ပါတို့ကို ရွေးချယ်နိုင်ပါတယ်-

- **Beta:** Beta distribution ဖြစ်ပါတယ်။ shape1 နဲ့ shape2 parameters တွေဖြစ်တဲ့ a နဲ့ b က အပေါင်းလက္ခဏာ (positive) ဖြစ်ရပါမယ်။ သူတို့ကို မသတ်မှတ်ထားရင် DISTRIBUTION ကနေ sample mean နဲ့ sample standard deviation ကိုသုံးပြီး ခန့်မှန်းပေးပါတယ်။ Observations အားလုံးက 0 နဲ့ 1 အပါအဝင် အဲဒီကြားမှာပဲ ရှိရပါမယ်။
- **Chi-square:** Chi-square distribution ဖြစ်ပါတယ်။ Degrees of freedom (df) ကို သတ်မှတ်ပေးရပါမယ်။ အနှုတ်လက္ခဏာ (Negative) ပါတဲ့ observations တွေကို လက်မခံပါဘူး။
- **Exponential:** Exponential distribution ဖြစ်ပါတယ်။ Scale parameter a ဟာ အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ Parameter ကို မသတ်မှတ်ထားရင် DISTRIBUTION ကနေ sample mean ကိုသုံးပြီး ခန့်မှန်းပေးပါတယ်။ အနှုတ်လက္ခဏာ ပါတဲ့ observations တွေကို လက်မခံပါဘူး။
- **Gamma:** Gamma distribution ဖြစ်ပါတယ်။ Shape နဲ့ scale parameters (a နဲ့ b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် DISTRIBUTION ကနေ sample mean နဲ့ sample standard deviation ကိုသုံးပြီး ခန့်မှန်းပေးပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။

- **Half-normal:** Half-normal distribution ဖြစ်ပါတယ်။ Data တွေကို location-free သို့မဟုတ် centralized ဖြစ်တယ်လို့ ယူဆထားပါတယ် (Location parameter=0)။ Scale parameter a ကို သင်ကိုယ်တိုင် သတ်မှတ်နိုင်သလို DISTRIBUTION ကို maximum likelihood method သုံးပြီး ခန့်မှန်းခိုင်းလို့လည်း ရပါတယ်။
- **Laplace:** Laplace ဒါမှမဟုတ် double exponential distribution ဖြစ်ပါတယ်။ LAPLACE ဟာ location နဲ့ scale parameter (a နဲ့ b) တွေကို အသုံးပြုပြီး၊ scale parameter (b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ Parameters တွေကို မသတ်မှတ်ထားရင် DISTRIBUTION ကနေ sample mean နဲ့ sample standard deviation ကိုသုံးပြီး ခန့်မှန်းပေးပါတယ်။
- **Logistic:** Logistic distribution ဖြစ်ပါတယ်။ LOGISTIC ဟာ location နဲ့ scale parameter (a နဲ့ b) တွေကို အသုံးပြုပြီး၊ scale parameter (b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ Parameters တွေကို မသတ်မှတ်ထားရင် DISTRIBUTION ကနေ sample mean နဲ့ sample standard deviation ကိုသုံးပြီး ခန့်မှန်းပေးပါတယ်။
- **Lognormal:** Lognormal distribution ဖြစ်ပါတယ်။ Scale နဲ့ shape parameters (a နဲ့ b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် DISTRIBUTION ကနေ sample data ရဲ့ natural logarithm ရဲ့ mean နဲ့ standard deviation ကိုသုံးပြီး ခန့်မှန်းပေးပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။
- **Normal:** Normal (ပုံမှန်) distribution ဖြစ်ပါတယ်။ Location parameter a ဟာ ဂဏန်းတန်ဖိုးတစ်ခုခုဖြစ်နိုင်ပြီး scale parameter b ကတော့ အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် DISTRIBUTION ကနေ sample mean နဲ့ sample standard deviation ကိုသုံးပြီး ခန့်မှန်းပေးပါတယ်။
- **Pareto:** Pareto distribution ဖြစ်ပါတယ်။ Threshold နဲ့ shape parameters (a နဲ့ b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် a ကို အငယ်ဆုံး observation နဲ့ ညီတယ်လို့ DISTRIBUTION က ယူဆပြီး b ကိုတော့ maximum likelihood method သုံးပြီး ခန့်မှန်းပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။
- **Student t:** Student's t distribution ဖြစ်ပါတယ်။ Degrees of freedom (df) ကို သတ်မှတ်ပေးရပါမယ်။
- **Uniform:** Uniform distribution ဖြစ်ပါတယ်။ UNIFORM မှာ minimum နဲ့ maximum parameter (a နဲ့ b) ပါဝင်ပြီး Parameter a ဟာ b ထက် ကြီးတာ ဒါမှမဟုတ် ညီနေရပါမယ်။ မသတ်မှတ်ထားရင် DISTRIBUTION က sample data ကနေ ယူဆတွက်ချက်ပေးပါတယ်။

- **Weibull:** Weibull distribution ဖြစ်ပါတယ်။ Scale နဲ့ shape parameters (a နဲ့ b) က အပေါင်းလက္ခဏာ ဖြစ်ရပါမယ်။ မသတ်မှတ်ထားရင် DISTRIBUTION က least square method သုံးပြီး ခန့်မှန်းပေးပါတယ်။ အနှုတ်လက္ခဏာ လက်မခံပါဘူး။

Distribution parameters (ဖြန့်ကျက်မှုဆိုင်ရာ ပါရာမီတာများ)

Distribution မဟာဗျူဟာနဲ့ parameter ရွေးချယ်စရာတွေကို ထောက်ပံ့ပေးပါတယ်။

- **Estimate from data:** ဒါကိုရွေးထားရင် data နဲ့ ရွေးချယ်ထားတဲ့ distribution အမျိုးအစားပေါ် မူတည်ပြီး distribution parameters တွေကို ခန့်မှန်းပေးပါတယ်။
- **Specify:** ဒါကိုရွေးထားရင် ရွေးချယ်ထားတဲ့ distribution အမျိုးအစားအတွက် parameters တွေကို သင်ကိုယ်တိုင် သတ်မှတ်ပေးနိုင်ပါတယ်။ (မှတ်ချက် - ရွေးချယ်နိုင်တဲ့ parameters တွေက ရွေးချယ်ထားတဲ့ distribution အမျိုးအစားပေါ် မူတည်ပြီး ကွဲပြားပါလိမ့်မယ်) ။

Transform (ပုံစံပြောင်းလဲခြင်း)

ပေးထားတဲ့ ရွေးချယ်စရာတွေက transform နဲ့ periodicity (အချိန်အပိုင်းအခြား) settings တွေကို သတ်မှတ်ပေးပါတယ်။

- **Natural log transform:** Data ထဲက ပြောင်းလဲနေတဲ့ amplitude (လွှဲကျယ်) တွေကို ဖယ်ရှားဖို့ သဘာဝ လော့ဂရစ်သမ် (natural logarithm - base e) ကိုသုံးပြီး data ကို ပုံစံပြောင်းပေးပါတယ်။
- **Standardize values:** Sequence သို့မဟုတ် time series variables တွေကို mean 0 နဲ့ standard deviation 1 ရှိတဲ့ sample တစ်ခုအနေနဲ့ ပြောင်းလဲပေးပါတယ်။
- **Difference:** မတည်ငြိမ်တဲ့ (non-stationary) variable တစ်ခုကို ကိန်းသေ mean နဲ့ variance ရှိတဲ့ တည်ငြိမ်တဲ့ (stationary) variable အဖြစ် ပြောင်းလဲဖို့ ဇယားမဆွဲခင် အသုံးပြုမယ့် differencing အတိုင်းအတာကို သတ်မှတ်ပေးပါတယ်။ အကွက်ထဲမှာ သင့်လျော်တဲ့ တန်ဖိုးကို ရိုက်ထည့်ပါ။
- **Seasonally difference:** တကယ်လို့ variable က ရာသီအလိုက် သို့မဟုတ် အချိန်အပိုင်းအခြားအလိုက် ပုံစံ (seasonal or periodic pattern) ပြနေရင်၊ ဇယားမဆွဲခင် variable ကို seasonally difference လုပ်ဖို့ ဒီ setting ကို သုံးနိုင်ပါတယ်။ (မှတ်ချက် - သတ်မှတ်ထားတဲ့ periodicity ရှိတဲ့ sequence သို့မဟုတ် time series variable တစ်ခုကို

quantitative variable တွေထဲက တစ်ခုအဖြစ် ရွေးချယ်ထားမှသာ ဒီ setting ကို အသုံးပြုနိုင်မှာပါ) ။

Proportion estimation formula (အချိုးအစား ခန့်မှန်းခြင်း ဖော်မြူလာ)

ပေးထားတဲ့ ရွေးချယ်စရာတွေက အချိုးအစားတွေကို ခန့်မှန်းဖို့ သုံးမယ့် ဖော်မြူလာကို သတ်မှတ်ပေးပါတယ်။

- **Blom's:** အချိုးအစား ခန့်မှန်းချက်တွေအပေါ် အခြေခံပြီး (r-3/8) / (w+1/4) ဖော်မြူလာကို သုံးကာ ranking variable အသစ်ကို ဖန်တီးပေးပါတယ်၊ အဲဒီမှာ w က case weights စုစုပေါင်းဖြစ်ပြီး r က rank (အဆင့်) ဖြစ်ပါတယ်။
- **Rankit:** (r-1/2) / w ဖော်မြူလာကို သုံးပါတယ်၊ w က လေ့လာတွေ့ရှိချက် (observations) အရေအတွက်ဖြစ်ပြီး r က 1 ကနေ w အထိ ရှိတဲ့ rank ဖြစ်ပါတယ်။
- **Tukey's:** (r-1/3) / (w+1/3) ဖော်မြူလာကို သုံးပါတယ်၊ r က rank ဖြစ်ပြီး w က case weights စုစုပေါင်း ဖြစ်ပါတယ်။
- **Van der Waerden's:** Van der Waerden ရဲ့ ပုံစံပြောင်းလဲခြင်းဖြစ်ပြီး r/(w+1) ဖော်မြူလာနဲ့ သတ်မှတ်ပါတယ်၊ အဲဒီမှာ w က case weights စုစုပေါင်းဖြစ်ပြီး r က 1 ကနေ w အထိ ရှိတဲ့ rank ဖြစ်ပါတယ်။
-

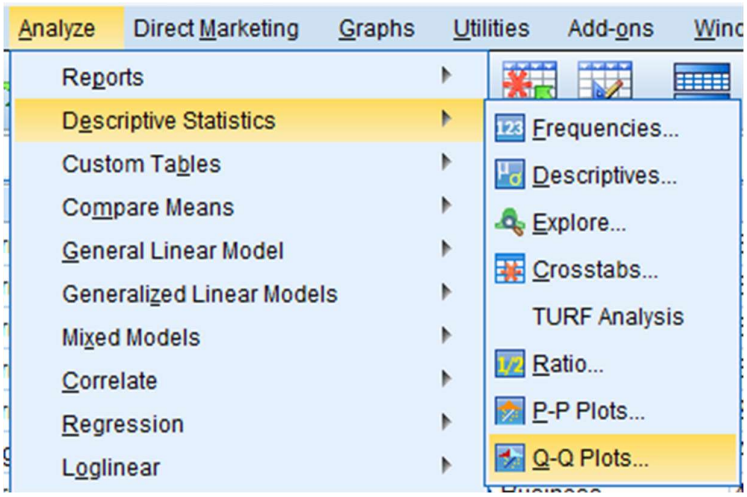
Rank assigned to ties (တန်ဖိုးတူညီမှုများအား အဆင့်သတ်မှတ်ခြင်း)

ပေးထားတဲ့ ရွေးချယ်စရာတွေက တန်ဖိုးတူညီနေတဲ့ (tie values) တွေကို ကိုင်တွယ်ဖြေရှင်းမယ့် နည်းလမ်းကို ထိန်းချုပ်ပေးပါတယ်။ အောက်ပါ ဇယား ၄ မှာ မတူညီတဲ့ နည်းလမ်းတွေက တန်ဖိုးတူညီနေတာတွေကို ဘယ်လို rank သတ်မှတ်ပေးသလဲဆိုတာ ပြသထားပါတယ်-

Value	Mean	Low	High	Break ties arbitrarily
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

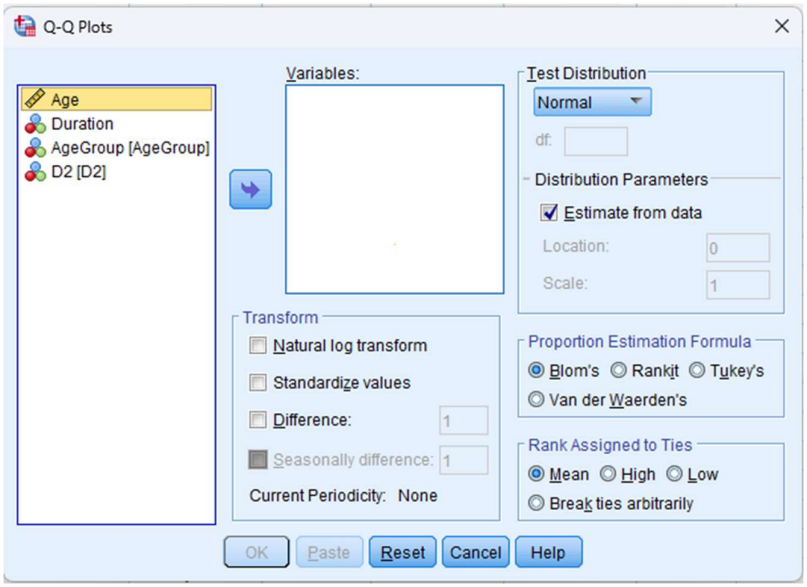
Obtaining Q-Q probability plots (Q-Q probability plots ရယူရန် အဆင့်များ)

၁။ Menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Descriptive Statistics > Q-Q Plots...



၂။ Numeric variables တစ်ခု သို့မဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ကာ Variables အကွက်ထဲကို ရွှေ့ထည့်ပါ။

၃။ Test distribution တစ်ခုကို ရွေးချယ်ပါ။



ရွေးချယ်နိုင်တဲ့ ထပ်ဆောင်းအချက်များ (Optionally) အနေနဲ့၊ ပုံစံပြောင်းထားတဲ့ တန်ဖိုးတွေ (transformed values) အတွက် probability plots တွေရယူဖို့ transformation options တွေကို ရွေးချယ်နိုင်ပြီး၊ မျှော်မှန်းထားတဲ့ ဖြန့်ကျက်မှု (expected distributions) တွေ တွက်ချက်ဖို့အတွက် နည်းလမ်းတစ်ခုကို သတ်မှတ်ပေးနိုင်ပါတယ်။

Reliability Analysis (ယုံကြည်စိတ်ချရမှု ခွဲခြမ်းစိတ်ဖြာခြင်း)

Reliability analysis ဟာ တိုင်းတာမှုစကေး (measurement scales) တွေရဲ့ ဂုဏ်သတ္တိတွေနဲ့ အဲဒီစကေးတွေကို ဖွဲ့စည်းထားတဲ့ အချက်အလက် (items) တွေကို လေ့လာခွင့်ပြုပါတယ်။ Reliability Analysis လုပ်ဆောင်ချက်က စကေးတွေရဲ့ ယုံကြည်စိတ်ချရမှုကို တိုင်းတာရာမှာ အများဆုံးအသုံးပြုလေ့ရှိတဲ့ အတိုင်းအတာ အတော်များများကို တွက်ချက်ပေးပြီး၊ စကေးထဲမှာပါဝင်တဲ့ item တစ်ခုချင်းစီကြားက ဆက်စပ်မှု အချက်အလက်တွေကိုလည်း ထောက်ပံ့ပေးပါတယ်။ အကဲဖြတ်သူများအကြား ယုံကြည်စိတ်ချရမှု (inter-rater reliability) ကို ခန့်မှန်းတွက်ချက်ဖို့အတွက် Intra-class correlation coefficients တွေကိုလည်း အသုံးပြုနိုင်ပါတယ်။

ဒါ့အပြင် အကဲဖြတ်သူ (raters) အမျိုးမျိုးကြားမှာ ယုံကြည်စိတ်ချရမှု ဘယ်လောက်ရှိသလဲဆိုတာကို ဆုံးဖြတ်ဖို့အတွက် အကဲဖြတ်သူများအကြား သဘောတူညီမှု (interrater agreement) ကို အကဲဖြတ်ပေးတဲ့ Fleiss' Multiple Rater Kappa စာရင်းအင်းများကိုလည်း ထောက်ပံ့ပေးပါတယ်။ သဘောတူညီမှု (agreement) ပိုများလေလေ၊ အကဲဖြတ်ချက်တွေဟာ အမှန်တကယ် အခြေအနေကို ထင်ဟပ်နေတယ်လို့ ပိုပြီး ယုံကြည်စိတ်ချရလေလေ ဖြစ်ပါတယ်။

Example (ဥပမာ)

ကိုယ့်ရဲ့ မေးခွန်းလွှာက ဖောက်သည်တွေရဲ့ ကျေနပ်မှုကို အသုံးဝင်တဲ့နည်းလမ်းနဲ့ တိုင်းတာနိုင်ရဲ့လား။ Reliability analysis ကို အသုံးပြုပြီး သင့်မေးခွန်းလွှာထဲက items တွေဟာ တစ်ခုနဲ့တစ်ခု ဘယ်လောက်အထိ ဆက်စပ်နေသလဲဆိုတာကို ဆုံးဖြတ်နိုင်ပါတယ်၊ စကေးတစ်ခုလုံးရဲ့ ထပ်တူညီမှု (repeatability) သို့မဟုတ် အတွင်းပိုင်း ကိုက်ညီမှု (internal consistency) ကို အလုံးစုံ ညွှန်းကိန်းတစ်ခုအနေနဲ့ ရယူနိုင်ပြီး၊ စကေးထဲကနေ ဖယ်ထုတ်သင့်တဲ့ ပြဿနာရှိတဲ့ items တွေကိုလည်း ခွဲခြားသိရှိနိုင်ပါတယ်။

Statistics (စာရင်းအင်းများ)

Variable တစ်ခုချင်းစီနဲ့ စကေးအတွက် Descriptives (ဖော်ပြချက်များ)၊ items တွေအားလုံးရဲ့ အကျဉ်းချုပ် စာရင်းအင်းများ (summary statistics)၊ inter-item correlations နဲ့ covariances တွေ၊ ယုံကြည်စိတ်ချရမှု ခန့်မှန်းချက်များ၊ ANOVA ဇယား၊ intraclass correlation coefficients၊ Hotelling's T^2၊ Tukey's test of additivity နဲ့ Fleiss' Multiple Rater Kappa တို့ ပါဝင်ပါတယ်။

Models (မော်ဒယ်များ)

အောက်ပါ ယုံကြည်စိတ်ချရမှုဆိုင်ရာ မော်ဒယ်တွေကို ရရှိနိုင်ပါတယ်-

- **Alpha (Cronbach):** ဒါက ပျမ်းမျှ inter-item correlation အပေါ်အခြေခံပြီး internal consistency ကို တိုင်းတာတဲ့ မော်ဒယ်ဖြစ်ပါတယ်။
- **Omega (McDonald's):** ဒီမော်ဒယ်က error covariances ပုံစံမျိုးနဲ့ ဒေသန္တရ item မှီခိုမှုမရှိတဲ့ (no local item dependence) factor တစ်ခုတည်းပါဝင်တဲ့ uni-dimensional ဖြစ်တယ်လို့ ယူဆပါတယ်။ မတူညီတဲ့ items နှစ်ခုရဲ့ covariance ဟာ သူတို့ရဲ့ loadings တွေ မြှောက်ထားတဲ့ ရလဒ်ဖြစ်တယ်လို့ ဆိုလိုပါတယ်။
- **Split-half:** စကေးကို နှစ်ပိုင်းခွဲလိုက်ပြီး အဲဒီ အပိုင်းနှစ်ခုကြားက ဆက်စပ်မှုကို စစ်ဆေးပေးတဲ့ မော်ဒယ်ဖြစ်ပါတယ်။
- **Guttman:** မှန်ကန်တဲ့ ယုံကြည်စိတ်ချရမှု (true reliability) အတွက် Guttman's lower bounds ကို တွက်ချက်ပေးပါတယ်။
- **Parallel:** Items အားလုံးမှာ တူညီတဲ့ variances တွေရှိပြီး၊ replications တွေတစ်လျှောက် တူညီတဲ့ error variances တွေရှိတယ်လို့ ယူဆတဲ့ မော်ဒယ်ဖြစ်ပါတယ်။
- **Strict parallel:** Parallel မော်ဒယ်ရဲ့ ယူဆချက်တွေကိုပဲ အသုံးပြုပြီး၊ items တွေတစ်လျှောက် တူညီတဲ့ means (ပျမ်းမျှကိန်းများ) ရှိတယ်လို့ ထပ်မံယူဆထားတဲ့ မော်ဒယ်ဖြစ်ပါတယ်။

Reliability Analysis Data Considerations (Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ)

- **Data:** Data တွေက dichotomous (နှစ်မျိုးခွဲ), ordinal (အစဉ်လိုက်), သို့မဟုတ် interval (ကြားကာလ) အမျိုးအစားတွေ ဖြစ်နိုင်ပေမယ့်၊ numeric (ဂဏန်း) အနေနဲ့ပဲ ကုဒ် (code) သတ်မှတ်ထားသင့်ပါတယ်။
- **Assumptions (ယူဆချက်များ):** လေ့လာတွေ့ရှိချက် (observations) တွေဟာ အမှီအခိုကင်း (independent) ဖြစ်ရမယ်၊ ပြီးတော့ errors တွေဟာ items တွေကြားမှာ ဆက်စပ်မှုမရှိ (uncorrelated) ဖြစ်ရပါမယ်။ Item စုံတွဲတိုင်းမှာ bivariate normal distribution ရှိသင့်ပါတယ်။ Item တစ်ခုချင်းစီဟာ စုစုပေါင်းရမှတ်နဲ့ မျဉ်းဖြောင့်ဆက်စပ်မှု (linearly related) ရှိစေဖို့ စကေးတွေဟာ ပေါင်းစပ်နိုင်သော (additive) ဖြစ်သင့်ပါတယ်။

  Fleiss' Multiple Rater Kappa statistics အတွက်ကတော့ အောက်ပါယူဆချက်တွေ သက်ရောက်မှုရှိပါတယ်-

  - မည်သည့် reliability statistic ကိုမဆို run ဖို့ rating variables အနည်းဆုံး နှစ်ခုကို ရွေးချယ်ထားရပါမယ်။
  - အကဲဖြတ်သူ (raters) တွေကြားမှာ အချိတ်အဆက် မရှိရပါဘူး။

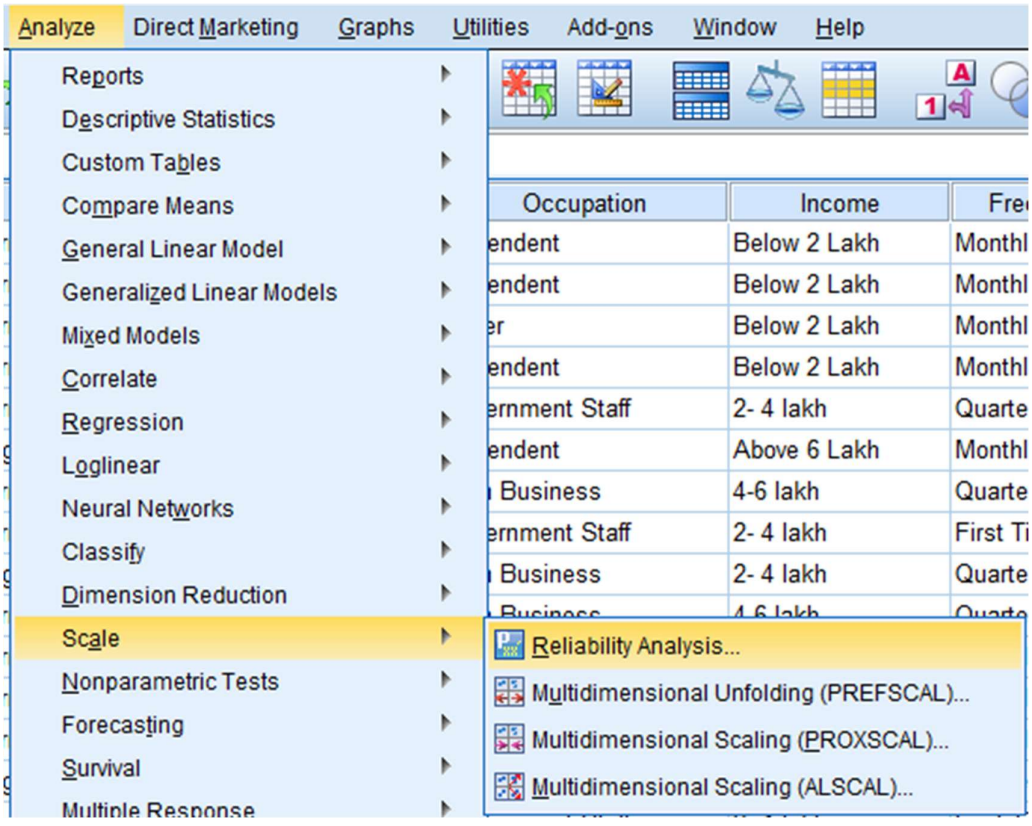
- အကဲဖြတ်သူ အရေအတွက်က ကိန်းသေ (constant) ဖြစ်ရပါမယ်။
- Subject တစ်ခုချင်းစီကို အကဲဖြတ်သူ တစ်ဦးတည်းပါဝင်တဲ့ တူညီတဲ့အုပ်စုကနေပဲ အကဲဖြတ်ရပါမယ်။
- သဘောထားကွဲလွဲမှုတွေအတွက် ဘယ်လိုအလေးချိန် (weights) ကိုမှ သတ်မှတ်ပေးလို့ မရပါဘူး။

- **Related procedures (ဆက်စပ်လုပ်ငန်းစဉ်များ)**

သင့် စကေး items တွေရဲ့ dimensionality ကို လေ့လာချင်ရင် (ပုံစံတူ အမှတ်ပေးမှုတွေကို ရှင်းပြဖို့ တည်ဆောက်မှု တစ်ခုထက်ပိုပြီး လိုအပ်သလား သိချင်ရင်) factor analysis ဒါမှမဟုတ် multidimensional scaling ကို အသုံးပြုပါ။ ဆင်တူတဲ့ variables အုပ်စုတွေကို ခွဲခြားသိရှိဖို့ဆိုရင်တော့ hierarchical cluster analysis ကိုသုံးပြီး variables တွေကို cluster (အုပ်စုဖွဲ့) လုပ်ပါ။

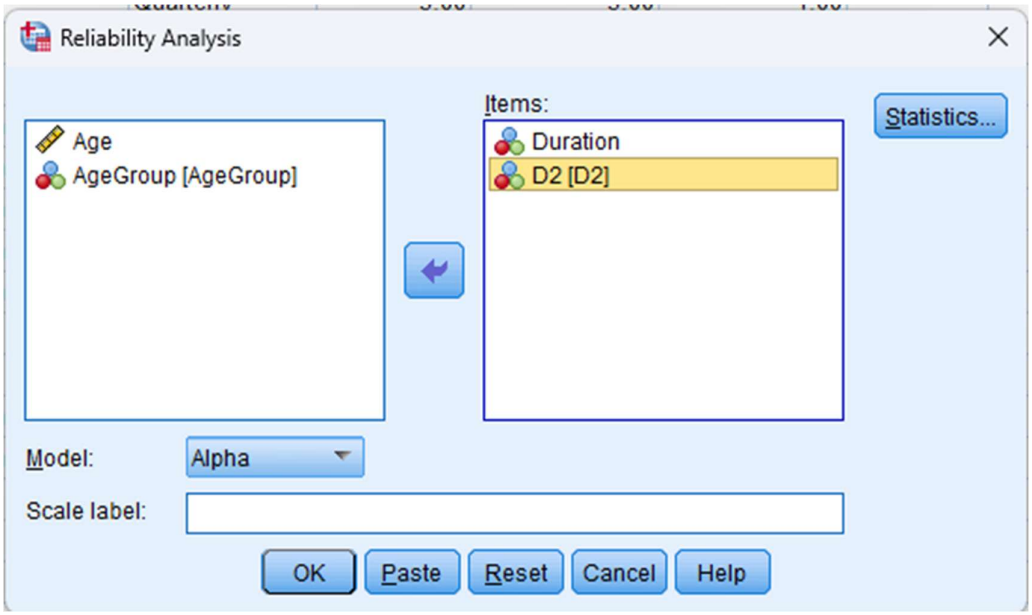
To obtain a Reliability Analysis (Reliability Analysis ရယူအသုံးပြုရန် အဆင့်များ)

၁။ Menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- Analyze > Scale > Reliability Analysis...



၂။ Additive scale တစ်ခုရဲ့ အစိတ်အပိုင်းတွေအဖြစ် ဖြစ်နိုင်ချေရှိတဲ့ variables နှစ်ခု ဒါမှမဟုတ် နှစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။

၃။ Model drop-down list ကနေ မော်ဒယ်တစ်ခုကို ရွေးချယ်ပါ။



၄။ လိုအပ်ပါက၊ သင့်ရဲ့ scale items တွေ ဒါမှမဟုတ် အကဲဖြတ်သူများအကြား သဘောတူညီမှုကို ဖော်ပြပေးတဲ့ စာရင်းအင်း အမျိုးမျိုးကို ရွေးချယ်ဖို့ **Statistics** ကို နှိပ်ပါ။

Reliability Analysis: Statistics (Reliability Analysis ဆိုင်ရာ စာရင်းအင်းများ)

သင့်စကေး၊ items တွေနဲ့ အကဲဖြတ်သူတွေကြား ယုံကြည်စိတ်ချရမှုကို ဆုံးဖြတ်ဖို့ interrater agreement တွေကို ဖော်ပြပေးတဲ့ စာရင်းအင်း အမျိုးမျိုးကို သင် ရွေးချယ်နိုင်ပါတယ်။ ပုံမှန် (default) အနေနဲ့ cases အရေအတွက်၊ items အရေအတွက်နဲ့ အောက်ပါ reliability estimates (ယုံကြည်စိတ်ချရမှု ခန့်မှန်းချက်များ) ကို ဖော်ပြပေးပါတယ်-

- **Alpha models:** Coefficient alpha ဖြစ်ပါတယ်၊ dichotomous data တွေအတွက်ဆိုရင်တော့ Kuder-Richardson 20 (KR20) coefficient နဲ့ ညီမျှပါတယ်။
- **Omega models:** ယုံကြည်စိတ်ချရမှုကို အကဲဖြတ်ဖို့ McDonald's omega ခန့်မှန်းချက်ပါ။
- **Split-half models:** ပုံစံနှစ်ခုကြားက ဆက်စပ်မှု (correlation between forms) ၊ Guttman split-half reliability၊ Spearman-Brown reliability နဲ့ တစ်ဝက်စီအတွက် coefficient alpha တို့ ပါဝင်ပါတယ်။
- **Guttman models:** Reliability coefficients lambda 1 မှ lambda 6 အထိ ပါဝင်ပါတယ်။

- **Parallel and Strict parallel models:** မော်ဒယ်ရဲ့ goodness of fit စမ်းသပ်ချက်၊ error variance, common variance နဲ့ true variance ခန့်မှန်းချက်တွေ၊ ခန့်မှန်း common inter-item correlation နဲ့ ယုံကြည်စိတ်ချရမှု ခန့်မှန်းချက်တွေ ပါဝင်ပါတယ်။
- **Descriptives for (ဖော်ပြချက်များ)** Scale တွေ သို့မဟုတ် cases တွေတစ်လျှောက်ရှိတဲ့ items တွေအတွက် descriptive statistics ကို ထုတ်ပေးပါတယ်။
    - **Item:** Cases တွေတစ်လျှောက်ရှိတဲ့ items တွေအတွက် descriptive statistics ကို ထုတ်ပေးပါတယ်။
    - **Scale:** စကေးတွေအတွက် descriptive statistics ကို ထုတ်ပေးပါတယ်။
    - **Scale if item deleted:** Item တစ်ခုချင်းစီကို အခြား items တွေနဲ့ ဖွဲ့စည်းထားတဲ့ စကေးနဲ့ နှိုင်းယှဉ်ပြတဲ့ summary statistics ကို ပြသပေးပါတယ်။ (တကယ်လို့ item ကိုသာ စကေးထဲကနေ ဖယ်ထုတ်လိုက်ရင် ရလာမယ့် scale mean နဲ့ variance၊ ဖယ်ထုတ်လိုက်တဲ့ item နဲ့ ကျန်တဲ့ items တွေကြားက correlation အစရှိတာတွေ ပါဝင်ပါတယ်) ။

Summaries (အကျဉ်းချုပ်များ)

စကေးထဲမှာပါဝင်တဲ့ items တွေအားလုံးရဲ့ item ဖြန့်ကျက်မှု (distributions) တွေနဲ့ပတ်သက်တဲ့ descriptive statistics ကို ထောက်ပံ့ပေးပါတယ်။

- **Means / Variances / Correlations / Covariances:** Items တွေရဲ့ ပျမ်းမျှကိန်းများ၊ ကွဲလွဲချက်များ၊ ဆက်စပ်မှုများနဲ့ covariances တွေအတွက် အကြီးဆုံး၊ အငယ်ဆုံး၊ ပျမ်းမျှတန်ဖိုးတွေ အပါအဝင် အချိုးအစားတွေကို တွက်ချက်ဖော်ပြပေးပါတယ်။

Inter-Item။ Items တွေကြားက correlations သို့မဟုတ် covariances တွေရဲ့ matrices ကို ထုတ်ပေးပါတယ်။

ANOVA Table။ ညီမျှတဲ့ means တွေ (equal means) ရှိ/မရှိ စမ်းသပ်ပေးပါတယ် (F test, Friedman chi-square, Cochran chi-square) ။

Interrater Agreement: Fleiss' Kappa (အကဲဖြတ်သူများအကြား သဘောတူညီမှု)

အကဲဖြတ်သူ (raters) တွေကြားမှာ ယုံကြည်စိတ်ချရမှု ဘယ်လောက်ရှိသလဲဆိုတာ ဆုံးဖြတ်ဖို့ အကဲဖြတ်သူတွေရဲ့ သဘောတူညီမှုကို အကဲဖြတ်ပေးပါတယ်။ သဘောတူညီမှု ပိုများလေလေ၊ အကဲဖြတ်ချက်တွေဟာ အမှန်တကယ် အခြေအနေကို ထင်ဟပ်နေတယ်လို့ ပိုပြီး ယုံကြည်စိတ်ချရလေလေ ဖြစ်ပါတယ်။ Generalized unweighted kappa statistic ဟာ အကဲဖြတ်သူ အရေအတွက် ကိန်းသေတစ်ခုအတွက် သဘောတူညီမှုကို တိုင်းတာပေးတာ ဖြစ်ပါတယ်။

Display agreement on individual categories:

သီးခြား category တစ်ခုချင်းစီအပေါ် သဘောတူညီချက်ကို ထုတ်ပြမလား ဆိုတာ သတ်မှတ်ပါတယ်။

Ignore string cases:

String variables တွေက case sensitive (အကြီးအသေး ခွဲခြားခြင်း) ဖြစ်မဖြစ် ထိန်းချုပ်ပေးပါတယ်။

String category labels are displayed in uppercase:

Output tables တွေထဲက category labels တွေကို အက္ခရာအကြီး (uppercase) နဲ့ ပြမလား၊ အသေး (lowercase) နဲ့ ပြမလား ထိန်းချုပ်ပါတယ်။

Missing (ပျောက်ဆုံးတန်ဖိုးများ)

User-missing နဲ့ system-missing တန်ဖိုးတွေကို ချန်လှပ်ထားမလား (Exclude both user-missing and system missing values) သို့မဟုတ် အကျုံးဝင်တဲ့ (valid) အချက်အလက်အဖြစ် သတ်မှတ်မလား (User-missing values are treated as valid) ဆိုတာကို ထိန်းချုပ်ပေးပါတယ်။

Intraclass correlation coefficient

Cases တွေအတွင်း တန်ဖိုးတွေရဲ့ ကိုက်ညီမှု (consistency) သို့မဟုတ် သဘောတူညီမှု (agreement) အတိုင်းအတာတွေကို ထုတ်ပေးပါတယ်။

- **Model:** Intraclass correlation coefficient တွက်ချက်ဖို့ မော်ဒယ် (Two-Way Mixed, Two-Way Random, One-Way Random) ကို ရွေးချယ်နိုင်ပါတယ်။
- **Type:** အညွှန်းကိန်း (index) အမျိုးအစားဖြစ်တဲ့ Consistency နဲ့ Absolute Agreement တို့ထဲက ရွေးချယ်နိုင်ပါတယ်။

RELIABILITY Command Additional Features

Command syntax language ကိုသုံးပြီး correlation matrix တွေကို ဖတ်တာ၊ ရေးတာ၊ split-half method အတွက် တခြားခွဲခြမ်းမှုတွေ သတ်မှတ်တာမျိုး လုပ်နိုင်ပါတယ်။

ROC Curves (ROC မျဉ်းကွေးများ)

ဒီလုပ်ဆောင်ချက်က subjects (လေ့လာခံရသူများ) ကို categories (အမျိုးအစား) နှစ်ခုရှိတဲ့ variable တစ်ခုနဲ့ ခွဲခြားသတ်မှတ်ပေးတဲ့ (classification schemes) တွေရဲ့ စွမ်းဆောင်ရည်ကို အကဲဖြတ်ဖို့အတွက် အသုံးဝင်တဲ့ နည်းလမ်းတစ်ခု ဖြစ်ပါတယ်။

ဥပမာ။

ဘဏ်တစ်ခုအနေနဲ့ ဖောက်သည်တွေကို ချေးငွေပြန်ဆပ်မယ့်သူတွေနဲ့ ချေးငွေမဆပ်ဘဲ ပျက်ကွက်မယ့်သူတွေဆိုပြီး မှန်မှန်ကန်ကန် ခွဲခြားသတ်မှတ်ဖို့က အရမ်းအရေးကြီးတဲ့အတွက်၊ ဒီလိုဆုံးဖြတ်ချက်တွေချနိုင်ဖို့ အထူးနည်းလမ်းတွေကို ဖန်တီးထားပါတယ်။ ဒီနည်းလမ်းတွေ ဘယ်လောက်အထိ ကောင်းကောင်း အလုပ်လုပ်သလဲဆိုတာကို အကဲဖြတ်ဖို့အတွက် ROC မျဉ်းကွေး (ROC curves) တွေကို အသုံးပြုနိုင်ပါတယ်။

- **Statistics (စာရင်းအင်းများ)။** ယုံကြည်မှု အတိုင်းအတာ (confidence interval) ပါဝင်တဲ့ ROC မျဉ်းကွေးအောက်က ဧရိယာ (Area under the ROC curve) နဲ့ ROC မျဉ်းကွေးရဲ့ ဆုံမှတ် (coordinate points) တွေကို ရရှိနိုင်ပါတယ်။
- **Plots (ပုံပြဇယားများ)။** ROC မျဉ်းကွေး (ROC curve) ကို ရရှိနိုင်ပါတယ်။
- **Methods (နည်းလမ်းများ)။** ROC မျဉ်းကွေးအောက်က ဧရိယာ ခန့်မှန်းချက်ကို nonparametric နည်းလမ်းနဲ့ ဖြစ်စေ၊ သို့မဟုတ် binegative exponential မော်ဒယ်ကို အသုံးပြုပြီး parametric နည်းလမ်းနဲ့ ဖြစ်စေ တွက်ချက်နိုင်ပါတယ်။

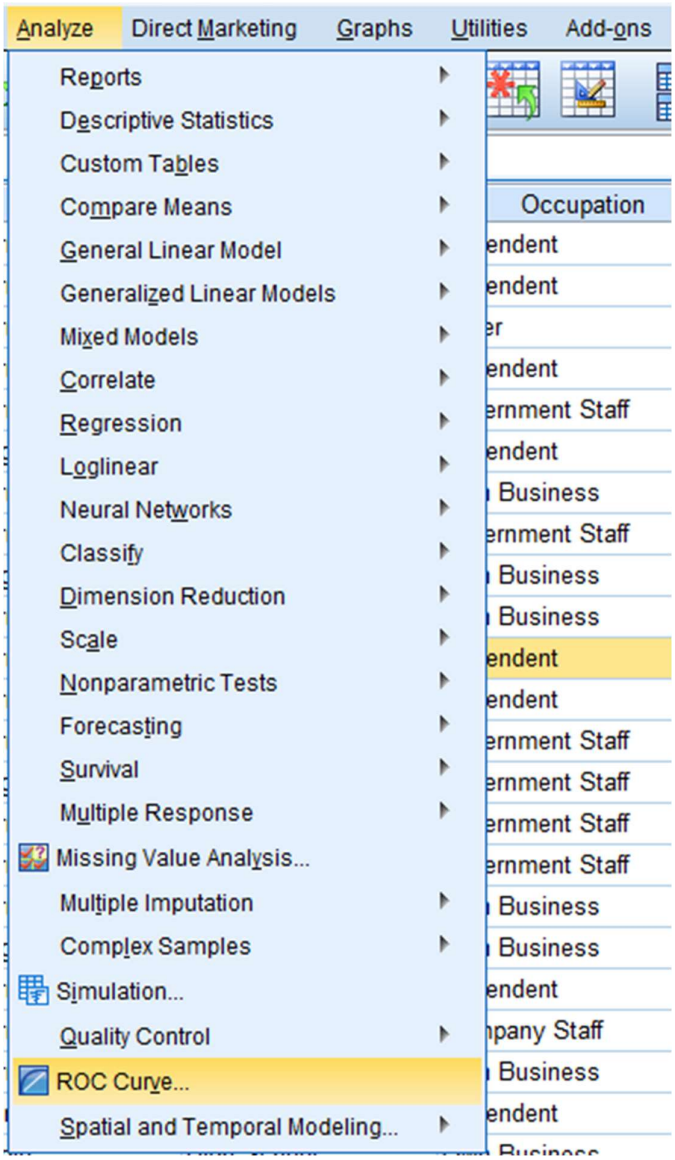
ROC Curve Data Considerations (ROC Curve အတွက် Data ဆိုင်ရာ ထည့်သွင်းစဉ်းစားရမယ့်အချက်များ)

- **Data (အချက်အလက်)။** စမ်းသပ်မယ့် variables (Test variables) တွေက အရေအတွက်ပြ (quantitative) ဖြစ်ရပါမယ်။ Test variables တွေဟာ များသောအားဖြင့် discriminant analysis သို့မဟုတ် logistic regression ကနေရလာတဲ့ ဖြစ်တန်စွမ်း (probabilities) တွေနဲ့ ဖွဲ့စည်းထားတတ်သလို၊ ဒါမှမဟုတ် subject တစ်ခုက category တစ်ခု ဒါမှမဟုတ် အခြား category တစ်ခုထဲကို ကျရောက်နေတယ်ဆိုတဲ့ အကဲဖြတ်သူရဲ့ "ယုံကြည်ချက် ပြင်းထန်မှု (strength of conviction)" ကို ညွှန်ပြတဲ့ ကျပန်းစကေး (arbitrary scale) ရမှတ်တွေနဲ့ ဖွဲ့စည်းထားတတ်ပါတယ်။ State variable ကတော့ ဘယ်အမျိုးအစားမဆို ဖြစ်နိုင်ပြီး၊ subject တစ်ခု အမှန်တကယ် ပါဝင်နေတဲ့ (true category) ကို ညွှန်ပြပါတယ်။ State variable ရဲ့ တန်ဖိုးက ဘယ် category ကို positive အဖြစ် သတ်မှတ်သင့်သလဲဆိုတာကို ညွှန်ပြပေးပါတယ်။

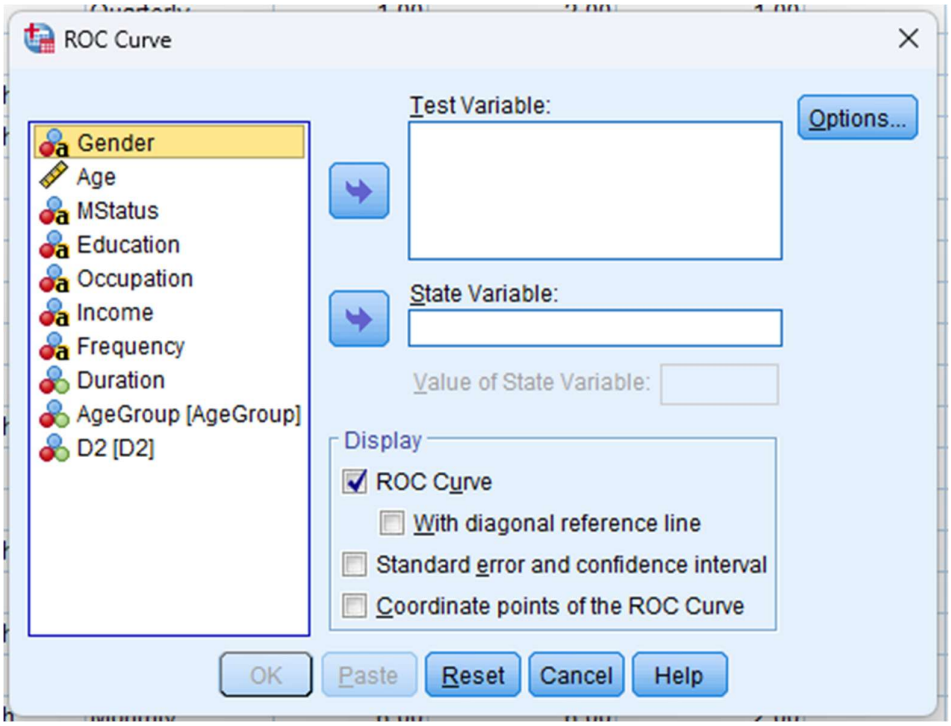
- **Assumptions (ယူဆချက်များ)။** အကဲဖြတ်သူရဲ့ စကေး (rater scale) ပေါ်က ဂဏန်းတွေ တိုးလာတာဟာ subject က category တစ်ခုထဲမှာ ပါဝင်တယ်ဆိုတဲ့ ယုံကြည်ချက် တိုးလာတာကို ကိုယ်စားပြုပြီး၊ စကေးပေါ်က ဂဏန်းတွေ လျော့ကျသွားတာကတော့ subject ဟာ အခြား category တစ်ခုထဲမှာ ပါဝင်တယ်ဆိုတဲ့ ယုံကြည်ချက် တိုးလာတာကို ကိုယ်စားပြုတယ်လို့ ယူဆထားပါတယ်။ အသုံးပြုသူအနေနဲ့ ဘယ်ဦးတည်ချက်က positive ဖြစ်တယ်ဆိုတာကို ရွေးချယ်ပေးရပါမယ်။ ဒါ့အပြင် subject တစ်ခုချင်းစီ အမှန်တကယ် ပါဝင်နေတဲ့ true category ကိုလည်း သိရှိထားပြီးဖြစ်တယ်လို့ ယူဆထားပါတယ်။

To Obtain an ROC Curve (ROC Curve ရယူအသုံးပြုရန် အဆင့်များ)

၁။ menus တွေထဲကနေ အောက်ပါအတိုင်း ရွေးချယ်ပါ- **Analyze > Classify > ROC Curve...**



၂။ စမ်းသပ်ဖြစ်တန်စွမ်း (test probability) variable တစ်ခု သို့မဟုတ် တစ်ခုထက်ပိုပြီး ရွေးချယ်ပါ။



၃။ State variable တစ်ခုကို ရွေးချယ်ပါ။

၄။ State variable အတွက် positive ဖြစ်မယ့် တန်ဖိုးကို သတ်မှတ်ပေး (identify) ပါ။

ROC Curve Options (ROC Curve ဆိုင်ရာ ရွေးချယ်စရာများ)

သင်ရဲ့ ROC analysis အတွက် အောက်ပါ ရွေးချယ်စရာတွေကို သတ်မှတ်ပေးနိုင်ပါတယ်-

- **Classification (အမျိုးအစားခွဲခြားခြင်း)။** Positive အဖြစ် ခွဲခြားသတ်မှတ်တဲ့အခါ ဖြတ်တောက်မယ့်တန်ဖိုး (cutoff value) ကို ထည့်သွင်းစဉ်းစားမလား (included) သို့မဟုတ် ချန်လှပ်ထားမလား (excluded) ဆိုတာကို သတ်မှတ်ပေးနိုင်ပါတယ်။ ဒီအပြင်အဆင်က လက်ရှိအချိန်မှာ output အပေါ် ဘာသက်ရောက်မှုမှ မရှိပါဘူး။
- **Test Direction (စမ်းသပ်မှု ဦးတည်ချက်) ။** Positive category နဲ့ ဆက်စပ်ပြီး စကေးရဲ့ ဦးတည်ချက်ကို သတ်မှတ်ပေးနိုင်ပါတယ်။
- **Parameters for Standard Error of Area (ဧရိယာ၏ စံအမှားအတွက် ပါရာမီတာများ) ။** မျဉ်းကွေးအောက်က ဧရိယာရဲ့ စံအမှား (standard error) ကို ခန့်မှန်းတွက်ချက်မယ့် နည်းလမ်းကို သတ်မှတ်ပေးနိုင်ပါတယ်။ ရရှိနိုင်တဲ့ နည်းလမ်းတွေကတော့ nonparametric

နဲ့ binegative exponential တို့ ဖြစ်ပါတယ်။ ဒါ့အပြင် confidence interval (ယုံကြည်မှု ကြားကာလ) အတွက် အဆင့်ကိုလည်း သတ်မှတ်ပေးနိုင်ပါတယ်။ သတ်မှတ်နိုင်တဲ့ အတိုင်းအတာ (range) ကတော့ 50.1% ကနေ 99.9% အတွင်း ဖြစ်ပါတယ်။

- **Missing Values (ပျောက်ဆုံးတန်ဖိုးများ)။** Missing values တွေကို ဘယ်လို ကိုင်တွယ်ဖြေရှင်းမလဲဆိုတာကို သတ်မှတ်ပေးနိုင်ပါတယ်။